با نام تو ای شط پر شوکت هر چه زیبای پاک

۲۷۱۳۸

۱۳۳۵ /۴/ ۲۰

دانشگاه شهید بهشتی

دانشکده علوم ریاضی - گروه آمار

پایان نامه

برای دریافت درجه کارشناسی ارشد در رشته آمار

موضوع:

معرفی روشهای بیزی گام به گام و تجربی و کاربرد آنها در نمونه گیری از جامعهٔ متناهی

استاد راهنما:

دکتر سیامک نوربلوچی

استاد مشاور:

دکتر محمد رضا مشکانی

نگارش:

مرضیه شکاری چاهستانی

دی ماه ۱۳۷۷

۲۷۱۳۸

36057

جویای راه خویش باش
از این سان که منم
در تکاپوی انسان شدن

در میان راه
دیدار می کنیم

حقیقت را

آزادی را

خود را
در میان راه

می بالد و به بار می نشیند
دوستی نی که توان مان می دهد
تا برای دیگران
مامنی باشیم و

یاوری
این است راه ما
راه تو
و من

مارگوت بیسکل

تقدیم به پدر و مادر عزیزم، آنها که در سفر پر هراس زندگی مرا آرامش بخشیدند

و

تقدیم به همسر عزیزم، او که چشمان شعله ورش، قندیل خاموش شبستانم را برافروخت

تشکر و قدردانی

- سپاس و ستایش، ایزد منان را که توفیق کسب قطره ای از دریای بیکران علم و دانش را به من عطا فرمود و یاریم نمود که رساله حاضر را انجام داده و به اتمام برسانم.
- برخود لازم می دانم که از استاد محترم و ارجمندم، جناب دکتر سیامک نوربلوچی که زحمت راهنمایی رساله حاضر را تقبل نموده و همواره مرا از راهنمائیهای ارزشمندشان بهره مند نمودند، کمال امتنان و تشکر را داشته باشم. از درگاه خالق یکتا، توفیق و سلامتی ایشان را خواستارم.

جا دارد از اساتید گرانمایه، آقایان دکتر مشکانی، دکتر شهلائی، دکتر شفیعی و دکتر ذکائی که زحمت داوری و مشاورت در ارائه هر چه بهتر این رساله را متحمل گشته اند، تشکر و قدردانی نمایم.

- و نیز، از اسطوره های صبر و ایثار و گذشت و مهربانی، پدر و مادر عزیز و گرامیم که الفبای زندگی را به من آموختند و همواره موفقیت خود را مدیون تلاش و زحمات آنان می دانم، صمیمانه سپاسگزارم و همت والایشان را ارج می نهم.
- همچنین از همسفر زندگیم که همواره یار و یاور و همراه و مشوق من بوده اند، به خاطر محبتهای بیدریغشان، صمیمانه سپاسگزارم.
- در پایان از دوست خوبم، خانم افسانه یزدانی به خاطر راهنماییها و زحمات بی‌شائبه، کمال تشکر را دارم.

فهرست مطالب

عنوان — شماره صفحه

چکیده فارسی

فصل اول – مقدمات نظریه تصمیم بیزی

۱-۱- مقدمه ۲

۱-۲- اصطلاحات و نمادهای مورد نیاز ۲

۱-۲-۱- چارچوب ۲

۱-۲-۲- جامعهٔ مورد نمونه گیری ۲

۱-۲-۳- تعریف نمونه ٤

۱-۲-٤- طرح نمونه گیری ٥

۱-۲-٥- طرح نمونه گیری بدون اطلاع ٥

۱-۲-٦- بهترین برآورد خطی نا اریب *(BLUE)* ٥

۱-۳- عناصر پایه ای نظریه تصمیم ٦

۱-۳-۱- تعریف تابع زیان ٦

۱-۳-۲- تعریف تابع مخاطره ٦

۱-۳-۳- تعریف بهتر بودن ٦

۱-۳-٤- قابلیت قبول ۷

۱-٤- مروری بر استنباط بیزی ۷

۱-٥- نحوهٔ برخورد بیزی ۸

۱-٦- قضایا و نتایج مورد نیاز ۱۱

۱-٦-۱- قضیه (لهمن ۱۹۸٥) ۱۱

۱-٦-۲- قضیه (لهمن ۱۹۸٥) ۱۱

عنوان | شماره صفحه

۳-٦-۱- قضیه (فرگسن ۱۹٦۷) ۱۲

٤-٦-۱- قضیه (اریکسون ۱۹٦۹) ۱۲

٥-٦-۱- قضیه (گلدستین ۱۹۷٥) ۱۲

٦-٦-۱- قضیه (سیرل ۱۹۷۱) ۱۳

۷-٦-۱- قضیه (براون ۱۹۸۱) ۱٤

۷-۱- نمونه گیری دو مرحله ای ۱٤

۱-۷-۱- مقدمه ۱٤

۲-۷-۱- تعاریف و اصطلاحات ۱٥

۸-۱- تحلیل بیزی ناپارامتری ۱٥

۱-۸-۱- انتخاب پیشین برای p (پارامتر ابر جامعه) ۱٦

۲-۸-۱- قضیه (ویلکس ۱۹٦۲) ۱۷

۳-۸-۱- قضیه (ویلکس ۱۹٦۲) ۱۷

فصل دوم- بیز گام به گام

۱-۲- مقدمه ۱۹

۲-۲- روش بیز گام به گام ۱۹

۳-۲- مشخصهٔ قابلیت قبول ۲٤

٤-۲- قابلیت قبول میانگین نمونه ۳۰

۱-٤-۲- قابلیت قبول و طرح ۳۰

۲-٤-۲- کاربرد بیز گام به گام در نمونه گیری از جامعهٔ متناهی ۳۲

٥-۲- ظرف پولیا ۳٦

عنوان | شماره صفحه

۶-۲- پسین پولیا ۳۷

۷-۲- میانگین نمونه برآورد بیز گام به گام برای میانگین جامعه متناهی ۴۱

فصل سوم- تعمیم پسین پولیا

۱-۳- مقدمه ۴۸

۲-۳- اطلاع پیشین ۴۸

۱-۲-۳- حدس پیشین برای جامعه ۴۸

۳-۳- کاربرد متغیر کمکی ۵۲

۱-۳-۳- برآوردگرهای نوع نسبتی ۵۲

۴-۳- انتخاب آزمایشها ۵۴

۱-۴-۳- مسئله کلی ۵۵

فصل چهارم- برآورد بیز تجربی

۱-۴- مقدمه ۵۹

۲-۴- برآوردگرهای بیز تجربی ۵۹

۳-۴- برآورد میانگین های طبقات ۶۱

۴-۴- برآورد نیرومند میانگین های طبقات ۷۲

۵-۴- نمونه گیری دو مرحله ای ۹۴

۶-۴- طرح گندم و جو در استان سمنان ۱۰۴

فهرست منابع ۱۲۱

چکیده انگلیسی

چکیده

رساله حاضر مشتمل بر چهار فصل است. در فصل اول، نمادها و تعاریف و قضایای مورد نیاز ارائه شده است. در فصل دوم، ابتدا یک مثال تکنیکی از روش بیز گام به گام ، توسط توزیع دو جمله ای شرح داده شده است. سپس با بیان قضیه ای و اثبات آن، رابطهٔ قابلیت قبول قاعده های تصمیم و بیز گام به گام بودن آنها را مشخص کرده ایم . براساس این یک برآوردگر، بیز گام به گام است، اگر دنباله ای متناهی از پیشینها موجود باشد، بطوریکه در مقابل همهٔ آنها بیز باشده براساس این قضیه تحت شرایطی یک برآوردگر، قابل قبول است اگر و تنها اگر بیز قدم بقدم باشد. قابلیت قبول میانگین نمونه، برآورد میانگین جامعهٔ متناهی را با استفاده از تکنیک بیز گام به گام و با نشان دادن اینکه میانگین نمونه، برآوردگر بیز گام به گام است، ثابت شده است. (البته قابل ذکر است که در سرتاسر این رساله، تابع زیان، توان دوم خطا در نظر گرفته شده است و فضای پارامتر و فضای نمونه، متناهی هستد.) در این اثبات، می بینیم که برای هر نقطهٔ نمونه ای داده شده، توزیع پیشین و یکتایی وجود ندارد، بنابراین توزیع های پسین (توزیع ندیده به شرط دیده) متفاوتی موجود است و در نتیجه برآورد بیز یکتا نیست،با توزیع ظرف پولیا نیز در این فصل آشنا می شویم و نشان داده می شود که توزیعهای پسین فوق، در حقیقت همان توزیع ظرف پولیا هستند و به عبارت دیگر معادل این است که واحدهای مشاهده شده را در ظرفی قرار دهیم و نمونه گیری پولیا را از ظرف انجام دهیم و بدین جهت آنها را پسین پولیا نامگذاری کرده اند. فرض شده است که در حقیقت پسینِ پولیا، یک روش بیز گام به گام است. در فصل سوم، روشهایی را برای تعمیم بحثهای فصل قبل ارائه داده ایم. در بعضی مواقع آمارشناس، مایل به پیدا کردن توزیع پیشین کاملی برای پارامتر مجهول γ در جامعه نیست، ولی مایل است حدسهایی را برای γ_1، ..... و γ_Nبدست آورد. بنابراین حالتهایی را در نظر گرفته ایم که حدسی پیشین برای کل جامعه یا برای هر عضو جامعه موجود است. در این فصل نشان داده ایم که چگونه می توان انواع مختلف اطلاع پیشین را در نمونه گیری از جامعه متناهی، بدون داشتن یک توزیع پیشین واقعی، ترکیب کرد و روشهای استنباطی را بصورت شیوه های بیزی و براساس توزیع پسین انجام داد.

در فصل ٤، روش بیز تجزیی را معرفی کرده ایم. در این فصل، برآورد بیز تجربی را برای میانگینهای طبقات با مدل بیزی نرمال و مدل نامشخص، بدست آورده ایم. سپس با دو روش مختلف، با توجه به پیشینهای دو مدل فوق، مخاطرهٔ بیز برآوردگرهای بیز تجربی میانگین طبقات را با برآوردگر کلاسیک آن مقایسه کرده ایم و در نهایت نتیجه گرفته ایم که روش بیز تجربی، مجانباً بهینه است. سپس تمام مراحل فوق را برای نمونه گیری دو مرحله ای نیز انجام داده ایم.

در خاتمه، روش بیزی تجربی فوق الذکر را روی مجموعه ای از داده های عملکرد گندم و جو در استان سمنان اعمال کرده ایم و برآورد متوسط عملکرد گندم و جو در این استان را به روش بیز تجربی محاسبه کرده‌ایم.

فصل اول

مقدمات نظریه تصمیم بیزی

۱-۱- مقدمه

در این فصل اصطلاحات و نمادهای مورد نیاز این رساله را ارائه می دهیم.

همچنین، مروری کوتاه بر استنباط بیزی شده است و عناصر پایه ای نظریه تصمیم و نظریه نمونه گیری بیزی از جامعه متناهی را نیز بطور خلاصه متذکر شده ایم. خواننده ای که با این مطالب آشناست می تواند از قرائت آن صرف نظر کرده و یا به اجمال صرفاً به جهت آشنایی با نمادگذاری بکار آمده در این رساله، آن را مرور کند.

۲-۱- اصطلاحات و نمادهای مورد نیاز

۱-۲-۱- چارچوب[1]

چنانچه بخواهیم از یک جامعهٔ متناهی نمونه بگیریم، می بایست نوعی فهرست کامل و دقیق و مشخص از اعضای آن جامعه داشته باشیم. بعنوان مثال اگر بخواهیم از مدارس تحت پوشش شبکه های بهداشت و درمان یک استان نمونه بگیریم، مسلماً برای اجرای یک نمونه گیری، داشتن فهرست کامل و دقیق شبکه ها و مدارس تحت پوشش آنها ضروری است. این فهرست جامع و مشخص را چارچوب می گویند. چارچوبها می توانند به صور مختلفی چون نقشه های جغرافیائی یا فایلهای کامپیوتری ساخته شوند.

۲-۱-۱- جامعهٔ مورد نمونه گیری[2]

جامعه ای است که بوسیله چارچوب در دسترس ما قرار می گیرد و آن را با V نمایش می‌دهیم. دو نظریه مختلف دربارهٔ نحوهٔ برخورد با مسأله نمونه گیری در مورد طرز تلقی ما از جامعه مورد نظر مطرح است:

1- Frame

2- Sampling population

۱- فرض جامعهٔ ثابت[1]: در این حالت هر عضو جامعه دارای یک عدد ثابت اما حقیقی و ناشناخته است که مقدار متغیر تحت بررسی است و در نتیجه جامعه، مجموعه ای یکتا از این اعداد ثابت نامعلوم است.

نمونه گیری کلاسیک با این فرض شروع می شود. در این شیوه، چارچوب نمونه گیری و تعیین دقیق جامعه مورد نمونه گیری ضروری است.

۲- فرض ابرجامعه[2]: در این فرض به هر عضو جامعه یک متغیر تصادفی وابسته است و به این لحاظ دارای ساختار تصادفی است. در نتیجه جامعهٔ ثابت وجود ندارد، بلکه جامعه مورد بررسی، خود متعلق به جامعهٔ بزرگتری است که اصطلاحاً آنرا اَبر جامعه می‌خوانیم.

در این نحوه صورت بندی، توجه به دو نکته زیر مهم است:

الف - نحوهٔ برخورد رایج و کلاسیک برفرض جامعهٔ ثابت متکی است.

ب - مسأله هائی که براساس مفاهیم پیش بینی، روشهای نظریه تصمیم، روشهای بیزی گسترش یافته‌اند، بر فرض ابر جامعه متکی هستند.

حال با توجه به این دو فرض که برای جامعه عنوان شد، علائم و نمادهائی را معرفی می کنیم:

جامعه متناهی، مجموعهٔ N عضو $(N < \infty)$است. بنابراین حجم جامعه را با N نشان می‌دهیم. عضوهای یک جامعه متناهی را قابل تشخیص یا متمایز می خوانیم، اگر بتوان به طور منحصر بفرد آنها را از 1 تا N بر چسب زد، به نحویکه برچسب یا شناسهٔ هر عضو معلوم باشد.

U یک مجموعهٔ عام[3] است و جامعه مورد نمونه گیری را تشکیل می دهد که آنرا با $U = \{U_1, U_2, \ldots, U_N\}$ نمایش می دهیم. واحد نمونه گیری[4]، $(i = 1,2,\ldots,N)$، U_i است.

1- Fixed population

2- Superpopulation approach

3- Universal set

4- Sampling unit

برای آسان کردن عملیات، بر چسب را به منزلهٔ واحد U_K مـی گـیریم. بنـابراین جامعـه را بـا $U = \{1,...,i,...,N\}$ نیز نمایش می دهیم. روشن است چنانچه تمایز وجود نداشته باشد، مجموعهٔ بالا معنی ندارد.

آن صفت یا صفات کمی یا کیفی را که تصمیم داریم،آنها را مورد بررسی قرار دهیـم را بـا y نمایش می دهیم. بنابراین متغیر مورد نمونه گیری، y می باشد و مقدار صفت y بـرای واحـد k ام، y_k است. مثلاً y_k میزان درآمد واحد Kام، به این معنی است که میزان درآمد فـرد U_K برابـر y_k اسـت که از فردی به فرد دیگر قابل تغییر است.

به طور کلی پارامتر جامعه متناهی را بصورت بردار زیر معرفی می کنند.

$$\underline{Y} = (y_1, y_2, ..., y_N) \qquad : \underline{Y} \in \Omega \subset R^N$$

در این رساله فرض می شود، فضـای پـارامتر $\underline{Y}$ زیـر مجموعـه ای متنـاهی از R^N اسـت. اگـر $b=(b_1,....,b_k)^T$ برداری k بعدی از اعداد حقیقی باشد، آنگاه منظور، از $Y(b)$ عبارتست از:

$$\underline{Y}(b) = \{y:\ y_i=b_j \qquad j=1,...,k \text{ و برای بعضی } i=1,...,N\}$$

توجه کنید $Y(b)$ کلیه نقاطی از فضای پارامتر است کـه در K مولفـه در نقطـه هـای b_i ثـابت گرفته شده اند.

برای فضای پارامتری $\underline{Y}$ و طرح p (که در ٤-٢-١ معرفی شده اسـت)، یـک نقطـهٔ نمونـه ای، بصورت زیر نوشته می شود:

$$z=(s,z_s) = (s,(z_{i1},....., z_{in(s)})^T)$$

٣-٢-١- تعریف نمونه

گیریم $s = \{i_1,...,i_j,...,i_{n(s)}\}$ بطوریکه برای $i = 1,...,N$ داریـم: $i_j \in U$،در اینصـورت دنبالهٔ s را یک نمونه (بدون جایگذاری و بدون حفظ ترتیب) می گوئیم. i_j ، jامیـن مولفـهٔ s و $n(s)$ تعداد مؤلفه های s یعنی حجم نمونه است. اگر $n(s)=n$ باشد، آنگاه حجم نمونه، ثابت است و اگـر

$n(s)=N$ باشد، عمل ما سرشماری یا تمام شماری است. $\mathcal{S}$ مجموعهٔ همه دنباله های s و یا بعبارت دیگر، کلیه نمونه های ممکن است.

۴-۲-۱- طرح نمونه گیری[1]

یک طرح نمونه گیری، براساس U عبارت است از زوج (S,p) که p توزیع احتمالی است که روی S تعریف می شود. بطوریکه داریم:

الف - برای هر $s \in S$ ، $p(s)>0$

ب - به ازای هر واحد که در جامعه U وجود دارد، لااقل یک نمونهٔ s وجود دارد که واحد مذکور را شامل می شود. تعریف بالا را در قالب ریاضی می توان آورد:

$$p: \quad \mathcal{S} \to [0,1]$$
$$\sum_{s \in S} p(s) = 1$$

۵-۲-۱- طرح نمونه گیری بدون اطلاع[2] از y

یک طرح نمونه گیری (S,p) بدون اطلاع است، اگر و فقط اگر تابع $p(.)$ به y_s بستگی نداشته باشد.

۶-۲-۱- بهترین برآورد خطی نا اریب

$\underline{a}'Y$، برای $\underline{\lambda}'\underline{\beta}$، بهترین برآورد خطی نااریب است اگر داشته باشیم،

$$1)\ E(\underline{a}'Y) = \underline{\lambda}'\underline{\beta} \qquad 2)\ Var(\underline{b}'Y) \geq Var(\underline{a}'Y)\ ;E(\underline{b}'Y) = \underline{\lambda}'\underline{\beta}$$

1- *Sampling design*

2- *noninformative*

۳-۱- عناصر پایه ای نظریه تصمیم

فرض می کنیم X متغیری تصادفی با تابع چگالی $f(X)$ نسبت به اندازه لبگ باشد. فضای نمونه را با $\mathcal{X}$، فضای پارامتر را با Θ و فضای اعمال را با A نمایش می دهیم. φ_x و φ_Θ بترتیب δ هیئت های مناسبی از $\mathcal{X}$ و Θ می باشند. از آنجا که مسائل مورد بررسی در این رساله، مسائل برآورد است، بنابراین همه جا فرض می کنیم که $A = \Theta$ است.

۱-۳-۱- تعریف تابع زیان[1]

تابع $L: \Theta \times \Theta \to R^+ \cup \{0\}$ را تابع زیان گوئیم، هر گاه $\varphi_\Theta \times \varphi_\Theta$ اندازه پذیر باشد.

در تمام قضایا و تعاریف مورد نیاز، تابع زیان، توان دوم خطا در نظر گرفته شده است.

۲-۳-۱- تعریف تابع مخاطره[2]

تابع ریسک یا مخاطرهٔ برآوردگر δ را با $R(\theta,\delta)$ نمایش می دهیم و بصورت زیر تعریف می کنیم:

$$R(\theta,\delta) = E_x[L(\theta,\delta(x))] = \int_x L(\theta,\delta(x))dF_x(x|\theta)$$

۳-۳-۱- تعریف بهتر بودن

برآوردگر δ_1 بهتر از برآوردگر δ_2 است اگر برای کلیه $\theta \in \Theta$ داشته باشیم:

$$R(\theta,\delta_1) \le R(\theta,\delta_2)$$

و نامساوی حداقل برای یک $\theta \in \Theta$ اکید باشد.

1- loss function

2- risk function