دانشگاه اصفهان

دانشکده علوم

گروه آمار

پایان‌نامه‌ی کارشناسی ارشد رشته‌ی آمار گرایش آماریاضی

معرفی توزیع‌های وزنی و کاربرد آنها در مشخصه‌سازی

استاد راهنما:

دکتر محمد حسین علامت‌ساز

پژوهشگر:

مهدی اسدی

مهرماه ۱۳۸۸

کلیه حقوق مادی مترتب بر نتایج مطالعات، ابتکارات و نوآوری‌های ناشی از تحقیق موضوع این پایان نامه متعلق به دانشگاه اصفهان است.

دانشگاه اصفهان
دانشکده علوم
گروه آمار

پایان‌نامه‌ی کارشناسی ارشد رشته‌ی آمارگرایش آمار ریاضی

مهدی اسدی

تحت عنوان

مطالعه ای بر توزیعهای وزنی و کاربرد آنها در مشخصه سازی

در تاریخ ۸۸/۷/ ۲۹ توسط هیأت داوران زیر بررسی با درجه بسیار خوب به تصویب نهایی رسید.

۱- استاد راهنمای پایان‌نامه دکتر محمد حسین علامت ساز با مرتبه‌ی علمی استاد امضاء

۲- استادداور داخل گروه پایان‌نامه دکتر مجید اسدی با مرتبه‌ی علمی دانشیار امضاء

۳- استاد داورخارج از گروه دکتر علی رجالی با مرتبه‌ی علمی دانشیار امضاء

امضای مدیر گروه

سپاسگزاری

سپاس خدای را که مرا شوق و توان ادراک آموخت.

مراتب سپاس و تشکر خود را از استاد گرانقدرم جناب آقای دکتر محمد حسین علامت ساز ابراز می‌دارم که در تمام مراحل تدوین این رساله با تلاش‌های بی وقفه و راهنمایی‌های ارزنده مرا یاری نمودند. همچنین از زحمات بی دریغ اساتید داور جناب آقای دکتر علی رجالی و جناب آقای دکتر مجید اسدی کمال تشکر را دارم.

از پدرم که سربلند زیستن و از مادرم که صبر و ایثار را به من آموخت نهایت تشکر و قدردانی را دارم.

تقدیم به

پدر و مادرم

به پاس تعبیر عظیم و انسانی شان از واژه ایثار

به پاس عاطفه سرشار و گرمای امید بخش وجودشان

و به پاس محبت های بی دریغ شان که هرگز فروکش نمی کند.

چکیده

وقتی یک محقق یک نمونه از مشاهدات را منطبق با یک مدل تصادفی خاص ثبت یا جمع آوری می‌کند مشاهدات ثبت شده به دلایل مختلف از قبیل غیر قابل مشاهده بودن برخی پیشامدها، تخریب جزئی مشاهدات و نمونه‌گیری با اطلاق شانس نابرابر به مشاهدات نمی‌توانند به عنوان یک نمونه تصادفی از توزیع اولیه جامعه مورد بررسی قرار گیرند. در این حالت‌ها یک توزیع مناسب و جایگزین برای مشاهدات ثبت شده توزیع موزون مرتبط با توزیع اولیه است. مفهوم توزیع‌های موزون به‌طور گسترده در مطالعات قابلیت اعتماد، آنالیز بقا، بوم شناسی، مطالعات حیات وحش و جوامع بشری مورد استفاده قرار می‌گیرد.

در این تحقیق بعضی مدل‌های کلی که منجر به توزیع‌های موزون می‌شوند را معرفی کرده و این توزیع‌ها را در حالت یک متغیره و چند متغیره بررسی می‌کنیم و به طور خاص به مطالعه توزیع‌های تعادلی و طول اریب می‌پردازیم. همچنین برخی نتایج ترتیب‌بندی تصادفی و حفظ کلاس‌های توزیع عمر تحت وزنی کردن را بررسی می‌کنیم. در نهایت از طریق اندازه‌های قابلیت اعتماد و با استفاده از توزیع‌های موزون مشخصه‌سازی‌های کلی از توابع احتمال به عمل آورده و همچنین از طریق رابطه آفین بین گشتاور اول توزیع اولیه و توزیع موزون مرتبط با آن مبحث مشخصه‌سازی خانواده نمایی طبیعی و سری توانی را دنبال خواهیم کرد.

هدف از انجام این تحقیق معرفی توزیع‌های موزون و آشنایی با برخی کاربردهای آنها می‌باشد. مشخصه‌سازی توزیع‌های احتمال با استفاده از ارتباط بین توزیع اولیه و توزیع موزون مرتبط با آن نیز از دیگر اهداف این پایان نامه است.

واژگان کلیدی: توزیع موزون، توزیع طول اریب، توزیع تعادلی، قابلیت اعتماد، ترتیب‌های تصادفی، کلاس‌های توزیع عمر، مشخصه‌سازی، خانواده نمایی طبیعی، خانواده سری توانی.

فهرست مطالب

عنوان **صفحه**

فصل اول: کلیات

۱-۱ مقدمه ۱

۱-۲ موضوع و هدف تحقیق ۱

۱-۳ اهمیت و کاربرد موضوع ۲

۱-۴ زمینه و تاریخچه موضوع تحقیق ۲

۱-۴ ساختار پایان‌نامه ۴

فصل دوم: توزیع‌های وزنی

۲-۱ مقدمه ۵

۲-۲ توزیع‌های وزنی یک متغیره ۶

۲-۲-۱ مدل‌های کلی منتهی به توزیع‌های وزنی ۷

۲-۲-۲ تابع وزن ۱۰

۲-۲-۳ توزیع طول اریب ۱۱

۲-۲-۴ توزیع تعادلی ۱۳

۲-۳ توزیع‌های وزنی چند متغیره ۱۴

۲-۳-۱ توزیع تعادلی چند متغیره ۱۷

۲-۴ اثر تابع وزن بر ساختار وابستگی ۱۸

۲-۴-۱ توزیع وزنی دو متغیره ۱۹

۲-۴-۲ وابستگی X_w و Y_w تحت وزن $w(x,y)=max(x,y)$ ۲۱

۲-۴-۳ وابستگی X_w و Y_w تحت وزن $w(x,y)=ax+by$ ۲۵

۲-۴-۴ نتایج کلی تاثیر تابع وزن بر وابستگی ۲۷

عنوان **صفحه**

۲-۴-۵ خلاصه و نتیجه‌گیری ۲۹

فصل سوم: حفظ کلاس‌های توزیع عمر و ترتیب‌های تصادفی تحت وزنی کردن

۳-۱ مقدمه ۳۱

۳-۲ حفظ کلاس‌های توزیع عمر تحت وزنی کردن ۳۲

۳-۲-۱ کلاس‌های توزیع عمر ۳۲

۳-۲-۲ توزیع‌های وزنی و منحنی لورنتس ۳۳

۳-۲-۳ اندازه‌های قابلیت اعتماد ۳۴

۳-۲-۳ توزیع‌های وزنی با نرخ شکست یکنوا ۳۵

۳-۲-۴ حفظ کلاس‌های *NBU (NWU)* و *IFRA (DFRA)* تحت وزنی کردن ۳۷

۳-۳ حفظ ترتیب‌های تصادفی تحت وزنی کردن ۴۲

۳-۳-۱ ترتیب‌های تصادفی ۴۲

۳-۳-۲ مقایسه بین متغیرهای تصادفی اصلی و وزنی ۴۳

۳-۴ حالات خاص ۵۱

۳-۴-۱ توزیع تعادلی ۵۲

۳-۴-۲ توزیع باقیمانده عمر ۵۳

۳-۵ برخی نتایج ترتیب‌های جزئی ۵۴

۳-۶ خلاصه و نتیجه‌گیری ۵۵

فصل چهارم: مشخصه‌سازی از طریق اندازه‌های قابلیت اعتماد به‌واسطه توزیع‌های وزنی

۴-۱ مقدمه ۵۶

۴-۲ مشخصه‌سازی با استفاده از روابط خطی بین اندازه‌های قابلیت اعتماد ۵۷

۴-۲-۱ ارتباط بین اندازه‌های قابلیت اعتماد و تابع توزیع ۵۷

عنوان **صفحه**

۴-۲-۲ قضایای مشخصه‌سازی با استفاده از روابط خطی ۵۸

۴-۲-۳ کاربردها ۶۲

۴-۳ مشخصه‌سازی با استفاده از توابع کلیدی از X و X_w ۶۳

۴-۳-۱ سیستم پیرسن توزیع‌ها ۶۳

۴-۳-۲ مشخصه‌سازی با استفاده از روابط بِین $r(.)$ و $m(.)$ ۶۳

۴-۳-۳ کاربردها ۶۵

۴-۴ مشخصه‌سازی با استفاده از اندازه‌های قابلیت اعتماد در حالت چند متغیره ۶۷

۴-۴-۱ اندازه‌های قابلیت اعتماد در حالت چند متغیره ۶۷

۴-۴-۲ قضایای مشخصه‌سازی ۶۹

۴-۵ خلاصه و نتیجه‌گیری ۷۰

فصل پنجم: مشخصه‌سازی با استفاده از توزیع‌های وزنی

۵-۱ مقدمه ۷۱

۵-۲ مشخصه‌سازی در خانواده نمایی طبیعی (NEF) ۷۳

۵-۲-۱ خانواده نمایی طبیعی (NEF) ۷۴

۵-۲-۲ رابطه آفین در حالت طول اریب ۷۵

۵-۳ مشخصه‌سازی در خانواده سری‌های توانی (PSF) ۸۴

۵-۳-۱ خانواده سری‌های توانی ۸۴

۵-۴ روابط آفین تصادفی ۸۹

۵-۶ خلاصه و نتیجه‌گیری ۱۰۰

منابع و مآخذ ۱۰۱

فهرست جدول‌ها

عنوان **صفحه**

جدول ۲-۱ توزیع‌های طول اریب مرتبط با بعضی توزیع‌های خاص ۱۲

جدول ۲-۲ توزیع‌های تعادلی مرتبط با بعضی توزیع‌های خاص ۱۴

جدول ۴-۱ مشخصه‌سازی از برخی توزیع‌ها در حالت طول اریب ۶۲

جدول ۴-۲ مشخصه‌سازی از برخی توزیع‌ها در حالت طول اریب مرتبه ۲ ..... ۶۲

جدول ۴-۳ مشخصه‌سازی از برخی توزیع‌ها وقتی $k_1 = E(X)$ و $k_2 = E(X_w)$ ۶۶

جدول ۴-۴ مشخصه‌سازی از برخی توزیع‌ها وقتی $k_1 = k_2 = 0$ ۶۶

مخففها

pdf	probability density function	تابع چگالی احتمال
pgf	probability generating function	تابع مولد احتمال
cgf	cumulant generating function	تابع مولد کومولان
LST	Laplas- Stieljes transformation	تبدیل لاپلاس استیلجس
NEF	Natural exponential family	خانواده نمایی طبیعی
PSF	Power series family	خانواده سری توانی
IFRA	Increasing failure rate average	متوسط نرخ شکست صعودی
DFRA	Decreasing failure rate average	متوسط نرخ شکست نزولی
IMRL	Increasing mean residual life	میانگین باقیمانده عمر صعودی
DMRL	Decreasing mean residual life	میانگین باقیمانده عمر نزولی
ILR	Increasing likelihood ratio	نسبت درستنمایی صعودی
IFR	Increasing failure rate	نرخ شکست صعودی
DFR	Decreasing failure rate	نرخ شکست نزولی
IRFR	Increasing reversed failure rate	نرخ شکست معکوس صعودی
DRFR	Decreasing reversed failure rate	نرخ شکست معکوس نزولی
PPS	Probability proportional to size	نمونه گیری متناسب با اندازه
NWU	New worse than used	نو بدتر از کهنه
NBU	New better than used	نو بهتر از کهنه
IVRL	Increasing variance residual life	واریانس باقیمانده عمر صعودی
DVRL	Decreasing variance residual life	واریانس باقیمانده عمر نزولی

فصل اول

کلیات

۱-۱ مقدمه

در این فصل ضمن بیان موضوع تحقیق، به ارائه هدف اصلی از انجام این تحقیق می‌پردازیم. در ادامه جهت آشنایی بیشتر خواننده با موضوع و بیان علل علاقه‌مندی و پرداختن به آن، اهمیت و کاربرد موضوع مورد نظر را بیان نموده و در این راستا زمینه و تاریخچه تحقیق را به تفصیل ذکر خواهیم کرد. لازم به ذکر است که در این پایان‌نامه از حروف بزرگ لاتین برای نمایش متغیر تصادفی و از پررنگ کردن حروف لاتین برای نمایش بردار استفاده می‌شود.

۱-۲ موضوع و هدف تحقیق

موضوع این پایان‌نامه معرفی توزیع‌های وزنی[1] و کاربرد آنها در مشخصه‌سازی می‌باشد. لذا با توجه به کاربرد وسیع توزیع‌های وزنی در مباحث مختلف آماری بخصوص استفاده از آنها در زمینه مشخصه‌سازی هدف از انجام این تحقیق این است که ضمن آشنا نمودن خواننده با مدل‌هایی که منجر به توزیع‌های وزنی می‌شوند، ساختار این

[1] Weighted distributions

توزیع‌ها را در حالات یک متغیره و چند متغیره بررسی نمائیم و سپس با توجه به ارتباط بین متغیر تصادفی اصلی و متغیر تصادفی وزنی مرتبط با آن، برخی توزیع‌های احتمال را مشخصه‌سازی خواهیم کرد.

۱-۳ اهمیت و کاربرد موضوع

در روش‌های آماری معمولاً یک نمونه تصادفی از جامعه انتخاب می‌شود و با استفاده از آن در مورد کل جامعه استنباط انجام می‌گیرد. به عبارتی نمونه تصادفی معرّف جامعه اصلی است. حال با توجه به این نکته که برخی عوامل از قبیل اریبی آشکار[1] ، نقص جزئی برخی مشاهدات ، نمونه گیری با تخصیص شانس نابرابر و... بر نحوه‌ی ثبت و جمع‌آوری مشاهدات مؤثر هستند، این سؤال مطرح است که آیا نمونه حاصل از توزیع اولیه مشاهدات، یعنی توزیع جامعه پیروی می‌کند؟ در این رساله نشان خواهیم داد که تحت چنین عواملی نمونه جمع‌آوری شده از یک توزیع وزنی مرتبط با توزیع اولیه پیروی می‌نماید.چون در اکثر تحقیقات آماری تنها ابزار قابل دسترس برای محقق همین نمونه‌ی تصادفی است. لذا مطالعه توزیع‌های وزنی، ارتباط آنها با توزیع اولیه و استفاده از این ارتباط در جهت مشخصه‌سازی توزیع اولیه مشاهدات، اهمیتی بیش از پیش خواهد یافت.

۱-٤ زمینه و تاریخچه موضوع تحقیق

متغیرها و توزیع های وزنی، اولین بار توسط فیشر[2](۱۹۳٤) مورد بررسی قرار گرفت. اما مطالعات سیستماتیک و فرم کنونی که برای توزیع‌های وزنی قائل هستیم توسط رائو[3](۱۹٦٥) ارائه شد. این مفهوم به طور گسترده در تجزیه و تحلیل داده‌های خانوار، مطالعات حیات وحش، قابلیت اعتماد، آنالیز بقا و زمینه‌های مختلف دیگر به‌کار برده می‌شود.

فرض کنید X یک متغیر تصادفی با تابع چگالی احتمال یا تابع جرم احتمال(pdf)، $f(x;q)$ باشد در حالی که q پارامتر مجهول توزیع مذکور است. همچنین فرض کنید احتمال ثبت یک مشاهده x از X برابر $w(x,b)$ باشد، به طوری‌که

$$P(\text{ثبت شدن} \mid X = x) = w(x, b)$$

[1] Visibility bias
[2] Fisher
[3] Rao

در اینجا تابع ثبت $w(x,b)$ یک تابع نامنفی با پارامتر b است که در اصل بیان کننده مکانیزم ثبت مشاهدات است. همان طور که ذکر شد واضح است که x ثبت شده یک مشاهده از X نیست بلکه مشاهده‌ای ازمتغیرتصادفی X_w است که دارای pdf زیر می‌باشد.

$$f_w(x;q,b)=\frac{w(x,b)f(x,q)}{w}$$

که در آن $w=E(w(X,b))$ یک عامل نرمال کننده است. متغیر تصادفی X_w متغیر تصادفی وزنی مرتبط با X و تابع وزن[1] $w(x,b)$ بوده و توزیع آن یک توزیع وزنی نامیده می‌شود. توزیع‌های وزنی برای مدل بندی روش‌های نمونه‌گیری که در آن احتمالات نمونه‌گیری متناسب با یک تابع وزن است، مطرح شدند. آنها همچنین برای تجزیه و تحلیل داده های واقعی، داده های زیست محیطی و نیز در نمونه‌گیری‌های هوایی مورد استفاده قرار می‌گیرند. در سال‌های اخیر با استفاده از توزیع‌های وزنی نتایج جالبی در خصوص مشخصه سازی توزیع‌های احتمال به‌دست آمده است. پاتیل[2] و رائو (۱۹۷۷)، بایاری و دگروت[3](۱۹۹۱) ضمن معرفی توزیع‌های وزنی به ارائه برخی کاربردهای آنها در نمونه‌گیری پرداختند. ناوارو[4] و همکاران (a-۲۰۰۱) نتایجی درخصوص برآورد پارامترها در نمونه گیری وزنی به دست آوردند. گوپتا[5] (۱۹۷۹) و موریس[6] (۱۹۸۲) با بررسی رابطه بین گشتاور اول متغیر تصادفی وزنی و متغیر اصلی مشخصه سازی‌هایی برای توزیع نمایی ارائه نمودند. جین[7] و همکاران (۱۹۸۹)، اسدی[8] (۱۹۹۸) و ناوارو و همکاران (b-۲۰۰۱) با استفاده از رابطه بین اندازه قابلیت اعتماد[9] متغیر تصادفی وزنی و متغیر اصلی مشخصه سازی‌هایی برای بعضی توزیع‌های احتمال خاص به‌دست آوردند.

پاکس[10] و همکاران (۲۰۰۳) و ناوارو و همکاران (۲۰۰۶) ضمن معرفی توزیع‌های وزنی در حالت چندمتغیره با استفاده از ارتباط بین اندازه قابلیت‌اعتماد متغیر تصادفی وزنی و متغیر اصلی، مشخصه‌سازی‌هایی در حالت

[1] Weight function
[2] Patil
[3] Bayarri and DeGroot
[4] Navarro
[5] Gupta
[6] Morris
[7] Jain
[8] Asadi
[9] Reliability measures
[10] Pakes

وزن‌های کلی‌تر به دست آوردند. همچنین با بررسی رابطه آفین[1] بین گشتاور اول متغیر تصادفی وزنی و متغیر اصلی و رابطه بین مجموع جزئی این دو متغیر مشخصه سازی‌هایی برای تابع توزیع F_X ارائه دادند.

۱-۵ ساختار پایان نامه

با توجه به اهداف تحقیق، این رساله در پنج فصل تدوین شده است.

فصل دوم شامل معرفی توزیع‌های وزنی در حالت یک متغیره و چند متغیره می‌باشد. همچنین در این فصل به بررسی اثرات تابع وزن بر ساختار وابستگی متغیرهای وزنی پرداخته شده است.

در فصل سوم مبحث حفظ و نگهداری ترتیب‌های تصادفی و کلاس‌های توزیع عمر تحت وزنی کردن را ارائه خواهیم نمود.

در فصل چهارم برخی روش‌های مشخصه‌سازی توزیع‌های احتمال با استفاده از اندازه‌های قابلیت اعتماد متغیر تصادفی اصلی و وزنی مطالعه خواهد شد.

نهایتاً در فصل پنجم مبحث مشخصه‌سازی توزیع‌های احتمال را با استفاده از رابطه آفین بین گشتاور اول متغیر تصادفی اصلی و متغیر وزنی مرتبط با آن و استفاده از رابطه آفین تصادفی بین این متغیرها دنبال خواهیم کرد.

[1] Affine relation

فصل دوم

توزیع‌های وزنی

۲-۱ مقدمه

وقتی یک محقق نمونه‌ای از مشاهدات را طبیعتاً منطبق با یک مدل خاص ثبت یا جمع آوری می‌کند به دلایل مختلف ممکن است نمونه‌ی جمع آوری شده از توزیع اولیه پیروی نکند. برخی از این دلایل به قرار زیر می باشند.

الف) غیر قابل مشاهده بودن پیشامدها

برخی از پیشامدها گرچه رخ می‌دهند اما ممکن است غیر قابل رؤیت باشند. بنابراین توزیع مشاهده شده توزیعی بریده شده است. یک مثال در این زمینه می‌تواند فراوانی خانواده‌هایی باشدکه والدین آنها آمادگی یا استعداد تولید بچه‌های زال را داشته باشند در حالی که هیچ فرزند زالی ندارند، هیچ گواه یا شاهدی وجود ندارد که خانواده‌ها مستعد باشند، مگر اینکه حداقل یک فرزند زال داشته باشند. لذا خانواده‌هایی با والدین مستعد ولی بدون فرزند زال، با خانواده‌های نرمال اشتباهاً در یک گروه قرار می‌گیرند. بنابراین فراوانی پیشامد صفر فرزند زال غیر قابل تحقق است.

ب) خرابی جزئی مشاهدات

مشاهداتی که به صورت طبیعی تولید می‌شوند، (تعدادتصادفات جاده‌ای) ممکن است به طور جزئی تخریب یا

تلف شوند، در چنین حالتی توزیع اولیه مخدوش خواهد شد و اگر مدل اساسی تخریب‌های جزئی توزیع اولیه(توزیع بقا) معلوم باشد، می‌توانیم توزیع مناسبی برای مشاهدات مورد بررسی ارائه نماییم.

پ) نمونه گیری با تخصیص شانس نابرابر به مشاهدات

یک روش مهم دیگر که در آن توزیع اولیه تغییر می‌کند یا اصلاح می‌شود. دادن شانس نابرابر به مشاهدات جهت نمونه‌گیری می‌باشد. این روش جزء لاینفک بعضی روش‌های نمونه‌گیری طبیعی و مناسب می‌باشد. برای مثال اگر بخواهیم توزیع فرزندان زال در خانواده‌هایی که مستعد تولید چنین فرزندانی هستند را مطالعه کنیم. می‌توانیم تعداد زیادی از خانواده‌ها را مورد مطالعه قرار دهیم و در هر خانواده تعداد فرزندان زال را تعیین کنیم. به این ترتیب خانواده‌هایی با حداقل یک فرزند زال یک توزیع دوجمله‌ای بریده شده تولید می‌کنند و احتمال زال بودن یک فرزند را از چنین توزیعی می‌توانیم برآورد کنیم. اما این روش تحقیق، اسرافکارانه است زیرا نسبت خانواده‌های غیر نرمال به طور کلی کوچک است و لذا تعداد خانواده‌هایی که در این زمینه به ما اطلاعات می‌دهند کم است، و محقق می‌بایست تعداد زیادی از خانواده‌ها را مورد مطالعه قرار دهد که این امر مستلزم صرف وقت و هزینه زیادی می‌باشد. لذا یک روش مناسب در چنین حالتی این است که ما یک فرزند زال را پیدا کرده و از طریق آن تعداد زال‌های خانواده‌اش را تعیین کنیم. اما در چنین روشی ممکن است شانس یکسان به همه‌ی خانواده‌های مورد بررسی داده نشود و شانس واقعی برای خانواده‌ای با x فرزند زال، ممکن است تابعی ازx باشد. یا به عنوان مثالی دیگر وقتی ما از طریق بیماران هموفیلی بستری شده در یک بیمارستان، خویشاوندان آنها را مورد مطالعه قرار می‌دهیم در واقع با مورد نمونه گیری با شانس نابرابر مواجه هستیم.

یکی از اهداف اصلی در هر تحقیق مشخصه سازی و تعیین چارچوب توزیع اولیه مشاهدات است. بنابراین مطالعه چگونگی اصلاح یا تغییر توزیع اولیه به واسطه دلایل ذکر شده در بالا از اهمیت بالایی برخوردار است. لذا در این فصل در بخش ۲ توزیع‌های وزنی و همچنین برخی مدل‌های کلی که منجر به توزیع‌های وزنی می‌شوند را در حالت یک متغیره بررسی نموده و در این راستا حالت‌های خاص توزیع طول اریب و تعادلی را تشریح می‌نمائیم، در بخش ۳ توزیع‌های وزنی را در حالت چند متغیره بیان نموده و در بخش ۴ به بررسی اثر تابع وزن بر ساختار وابستگی در توزیع‌های وزنی دو بعدی، خواهیم پرداخت.

۲-۲ توزیع‌های وزنی یک متغیره

فرض کنید X یک متغیر تصادفی با تابع چگالی $f(x)$ باشد. همچنین فرض کنید احتمال ثبت یک مشاهده‌ی x از X متناسب با تابع نامنفی از اندازه x مانند $w(x)$ باشد. در حالت کلی الزامی برای اینکه تابع نـامنفی w از

جنس احتمال باشد، یعنی دارای کران بالای یک باشد، وجود ندارد. لذا اگر $w : R \to R^+$ تابعی باشد به طوری که $0 < E\left(w\left(X\right)\right) < \infty$ آنگاه

$$F_w\left(x\right) = \frac{1}{E\left(w\left(X\right)\right)} \int_{-\infty}^{x} w\left(u\right) dF\left(u\right) \qquad (۱-۲)$$

تابع توزیع وزنی مرتبط با F و تابع وزن w بوده و تابع چگالی آن به صورت زیر نمایش داده خواهد شد.

$$f_w\left(x\right) = w\left(x\right) f\left(x\right) / E\left(w\left(X\right)\right) \qquad (۲-۲)$$

در ادامه جهت آشنایی بیشتر با توزیع های وزنی، مدل‌های کلی که منجر به توزیع های وزنی با توابع وزن مختلف می‌شوند را ارائه می‌دهیم.

۲-۲-۱ مدل‌های کلی منتهی به توزیع‌های وزنی

الف) احتمالات نمونه گیری در بررسی‌های نمونه‌ای

در این مدل جهت بهبود کارایی برآوردگرهای پارامترهای مجهول، به واسطه طرح ارائه شده برای انتخاب نمونه، تابع توزیع اولیه متغیر تصادفی تغییر پیدا می کند. یک مثال شناخته شده در این زمینه یک روش نمونه گیری است که در متدلوژی بررسی نمونه، نمونه گیری متناسب با اندازه (PPS[1]) نامیده می شود.

فرض کنید Y متغیرتصادفی کمکی یا همراه متغیرتصادفی اصلی تحت مطالعه X باشد و $f_1\left(y\right)$ ، $f_2\left(x \mid y\right)$ $f_3\left(x\right)$ و $f_4\left(y \mid x\right)$ به ترتیب چگالی‌های حاشیه ای Y, شرطیX به شرط Y , حاشیه‌ای X , شرطیY به شرط X باشند . حال چنانچه در ابتدا y را به عنوان یک تحقق از متغیر تصادفی Y_w با تابع چگالی زیر مشاهده کنیم

$$\frac{g\left(y\right).f_1\left(y\right)}{\int g\left(y\right) f_1\left(y\right) dy}$$

که در آن $g(y)$ یک تابع نامنفی است که جهت ماکزیمم کردن کارایی برآوردگرهای مجهول جامعه مورد استفاده قرار گرفته است، آنگاه با استفاده از توزیع شرطی $f_2\left(x \mid y\right)$ یک مشاهده ازX ساخته می‌شود و لذا تابع چگالی احتمال متغیر تصادفی X_w به‌صورت زیر خواهد بود

[1] Probability proportional to size

$$\frac{\int f_2(x\mid y)g(y)f_1(y)dy}{\int g(y)f_1(y)dy}=f_3(x)w(x)\Big/\int f_3(x)w(x)dx$$

که در آن $w(x)=\int f_4(y\mid x)g(y)dy$ و این توزیع یک توزیع وزنی از X با تابع وزن $w(x)$ است که لزوماً دارای کران بالای یک نیست .

ب) مدل زیان (رائو،۱۹۶۵)

فرض کنید بخواهیم یک نمونه از متغیرتصادفی X با چگالی،$f(x)$ تهیه کنیم، ولی مشاهده z تحت تاثیر یک فرآیند زیان قرار گرفته و در نهایت ما یک مشاهده از توزیع شرطی با چگالی ، $c(z\mid X=x)$ را خواهیم داشت و چگالی حاشیه‌ای Z نیز به صورت زیر خواهد بود

$$f_Z(z)=\int c(z\mid x)f(x)dx$$

که صورتی از یک توزیع آمیخته است.

حال اگر فرض کنیم یک مشاهده زمانی ثبت می‌شود که مقدار اولیه به واسطه فرآیند زیان مذکور بدون تغییر باقی بماند، آنگاه مشاهدات ثبت شده از متغیر تصادفی X_w با چگالی زیر خواهند بود.

$$f_w(x)=\frac{c(x\mid x)f(x)}{E\left(c(X\mid X)\right)}$$

که یک توزیع وزنی با تابع وزن $w(x)=c(x\mid x)=c(Z=x\mid X=x)$ خواهد بود. تابع وزن در اینجا احتمال شرطی، پیشامد بدون تغییر ماندن مشاهده x تحت فرآیند زیان می باشد.

به عنوان مثال یک توزیع بریده شده یک مورد خاص از این مدل است که در آن $c(x\mid x)$ در یک ناحیه معین از فضای متغیر X مقدار صفر را اختیار کرده و در مکمل آن مقدار یک را می‌گیرد.

پ) اریبی آشکار

اریبی آشکار مشکلی است که در تکنیک‌های سرشماری(بویژه در سرشماری‌های هوایی برای برآورد چگالی جامعه حیات وحش) با آن مواجه هستیم. علت اصلی عدم دقت سرشماری هوایی عموماً به دلیل همین اریبی آشکار است که به عوامل مختلف محیطی از جمله پوشش برف، آب و هوا، خستگی مشاهده کننده و ناحیه عملیاتی بستگی دارد. این عوامل می‌توانند سبب اریبی برآورد چگالی جامعه باشند. فرض کنید یک متغیر تصادفی گسسته X با چگالی $f(x)$ را بررسی کنیم . به عنوان مثال X ممکن است تعداد افراد در یک گروه یا گردایه باشد. در چنین حالتی $f(x)$ احتمال این پیشامد است که گروه شامل x فرداست. حال فرض کنید یک گروه تنها زمانی ثبت می‌شود که حداقل یکی از افراد آن، رؤیت شود و هر شخص برای رؤیت شدن دارای یک

شانس مستقل b باشد، آنگاه احتمال اینکه یک گروه مشاهده شده، دارای x عضو باشد برابر است با

$$f_w(x)=\frac{w(x)f(x)}{E(w(X))} \qquad (۳-۲)$$

که در آن $w(x)=[1-(1-b)^x]$.

حد $f_w(x)$ وقتی $b\to 0$ برابر است با $xf(x)/E(X)$. لذا این عبارت زمانی که b خیلی کوچک است تقریب خوبی برای رابطه (۲-۳) خواهد بود.

معنی داری تابع وزن $w(x)=1-(1-b)^x$ با استفاده از مثال‌های زیر آشکار خواهد شد.

مثال ۲-۱ :(هالدان[1] ۱۹۳۸, فیشر ۱۹۳۴ , رائو۱۹۶۵, نیل و اسکول[2] ۱۹۶۶)

اگر ما بخواهیم توزیع تعداد فرزندان زال X را در خانواده‌هایی که مستعد تولید چنین فرزندانی هستند را مطالعه کنیم یک روش نمونه گیری مناسب این است که ابتدا یک فرزند زال را پیدا نموده و از طریق آن تعداد فرزندان زال X_w خانواده او را تعیین کنیم . اگر احتمال یافتن یک زال b باشد آنگاه احتمال این‌که یک خانواده با X فرزند زال ثبت شود، برابر $w(x)=1-(1-b)^x$ خواهد بود و در چنین حالتی X_w دارای توزیع دوجمله‌ای وزنی با تابع وزن فوق خواهد بود .

مثال ۲-۲ :(کوک و مارتین[3] ۱۹۷۴)

فرض کنید در سرشماری هوایی جامعه حیات وحش،حیوانات به صورت گروهی یافت می‌شوند و تعداد اعضای گروه X دارای چگالی $f(x)$ بوده و احتمال رؤیت یک حیوان برابر b باشد. غیر قابل رؤیت بودن در اینجا یکی از دلایل اصلی عدم مشاهده برخی حیوانات است. حال چنانچه به محض مشاهده حداقل یک حیوان یک شمارش کامل از گروه به عمل آید و تعداد حیوانات آن گروه ثبت شود. آنگاه احتمال ثبت یک گروه با x حیوان برابر $w(x)=1-(1-b)^x$ بوده و تعداد اعضای گروه مشاهده شده X_w دارای توزیع وزنی با تابع وزن $w(x)$ خواهد بود.

ت) نمونه گیری دو مرحله‌ای (یک قضیه حدی)

ابتدا یک متغیر تصادفی X را در نظر می گیریم طوری که

$$P(X=i)=p_i \quad , \quad i=1,2,...$$

[1] Haldane
[2] Neel and Schull
[3] Cook and Martin