بسم الله الرحمن الرحیم

دانشگاه اصفهان

دانشکده علوم

گروه آمار

پایان نامه‌ی کارشناسی ارشد رشته‌ی آمار گرایش اقتصادی - اجتماعی

خانواده جدیدی از توزیع های گوسی وارون

استاد راهنما:

دکتر محمد حسین علامت ساز

پژوهشگر:

زهرا زائری

بهمن ماه ۱۳۸۹

دانشگاه اصفهان

دانشکده علوم

گروه آمار

پایان نامه ی کارشناسی ارشد رشته‌ی آمار گرایش اقتصادی - اجتماعی

خانم زهرا زائری امیرانی

تحت عنوان

خانواده جدیدی از توزیع های گوسی وارون

در تاریخ ۸۹/۱۱/۲۰ توسط هیأت داوران زیر بررسی و با درجه بسیار خوب به تصویب نهایی رسید.

۱- استاد راهنمای اول پایان نامه دکتر محمد حسین علامت ساز با مرتبه ی علمی استاد امضا

۲- استاد داور داخل پایان نامه دکتر افشین پرورده با مرتبه ی علمی استادیار امضا

۳- استاد داور خارج از گروه دکترعلی رجالی با مرتبه ی علمی دانشیار امضا

امضای مدیر گروه

تشکر و قدردانی:

حمد و سپاس بیکران خداوند سبحان را سزاست که این بنده را بر اتمام این تحقیق توانا ساخت. به مصداق کلام زیبای «من لم یشکر المخلوق لم یشکر الخالق » بر خود لازم می دانم از کلیه بزرگوارانی که مرا در این تحقیق یاری نمودند تشکر و قدردانی نمایم.

- جناب آقای **دکترمحمد حسین علامت ساز،** استاد محترم راهنما که از رهنمودهای بیدریغشان در تمام مراحل پژوهش بهره مند بوده ام.

- اساتید محترم داور و ناظر که نظرات ارزشمندشان بر غنای این پژوهش افزود.

و

- خانواده عزیزم که کمک های بیدریغشان پیمودن این راه را بر من هموار ساخت.

و

- همسرعزیزم که با حوصله تمام مرا در این امر یاری رساند.

تقدیم به:

پدر و مادر و همسرم

چکیده

خانواده توزیع های گوسی وارون از توزیع های نامتقارنی تشکیل شده است که به میزان قابل توجهی مشابه خانواده توزیع‌های گوسی در ارائه راه‌حل‌های ساده استنباطی در مدل سازی و تحلیل داده‌های مثبت چوله مورد استفاده قرار می‌گیرند. این توزیع به‌ویژه می‌تواند به‌عنوان مدلی برای طول عمر خرابی و بررسی قابلیت اعتماد یک سیستم با فرآورده های صنعتی به‌کار رود.

در این پایان نامه، ابتدا خانواده توزیع‌های گوسی وارون و ویژگی‌های اصلی توزیعی و استنباطی آن مورد بررسی قرار داده می‌شود و توزیع‌های مرتبط با این خانواده از قبیل توزیع‌های نمونه‌گیری و گوسی وارون چند متغیره معرفی خواهند شد. سپس توزیع گوسی وارون تعمیم یافته و توزیع های $\alpha-$ گوسی وارون مورد بررسی قرارداده می‌شوند و خصوصیات آن‌ها معین می‌گردند. همچنین یک مدل جدید توزیع گوسی وارون معروف به توزیع گوسی نوع وارون را تعریف می‌شود و خصوصیات توزیعی و استنباطی آن‌مطالعه و کاربرد آن در یک مجموعه داده نشان داده می‌شود. بالاخره، توزیع گوسی وارون آمیخته مورد بررسی قرارمی‌گیرد و روش جدیدی برای محاسبه گشتاوهای توزیع ارائه می‌شود. سپس از آن برای انتخاب مدل استفاده می‌شود و کاربرد توزیع گوسی آمیخته با استفاده از چهار مجموعه داده تشریح خواهد شد.

کلید واژه ها: توزیع گوسی وارون، توزیع گوسی وارون تعمیم یافته، توزیع گوسی وارون چند متغیره، توزیع لاگرانژ چند متغیره، روش درستنمایی.

فهرست مطالب

عنوان **صفحه**

عنوان **صفحه**

فصل سوم: توزیع گوسی وارون تعمیم یافته

فصل چهارم: مدل جدید توریع گوسی وارون

عنوان **صفحه**

فصل پنجم: **توریع گوسی وارون آمیخته**

فهرست جدول ها

عنوان | **صفحه**

فهرست نمودارها

عنوان **صفحه**

فصل اول

معرفی توزیع گوسی وارون

شرحی بر موضوع

خانواده گوسی وارون[1] (IG)از توزیع‌های نامتقارنی تشکیل شده‌است که به طور گسترده برای مدل‌سازی و تحلیل داده‌های نامنفی چوله مورد استفاده قرار می‌گیرد. این توزیع ها به ویژه به عنوان مدلی برای طول عمر خرابی وبررسی قابلیت اعتماد یک فرآورده صنعتی وهمچنین برای توصیف فراوانی سرعت باد جهت ارزیابی انرژی باد مورد استفاده قرار می‌گیرند. توزیع گوسی وارون توسط شرودینگر[2] (۱۹۱۵) به دلیل خواص مهم فیزیکی آن تحت عنوان توزیع زمان اولین گذر در حرکت براونی مطرح گردید. وی مشاهده نمود که اگر $X(t)$ یک حرکت براونی با $X(0)=0$ و رانش مثبت δ و ثابت انتشار β^2 باشد، آنگاه زمان اولین گذر از موقعیت مکانی $a>0$ دارای توزیع گوسی وارون با پارامترهای $\mu=\frac{a}{\mu}, \lambda=\frac{a^2}{\beta^2}$ است.

---

[1] Inverse Gaussian
[2] Schrodinger

در مقاله ای مروری فولکز و چیکارا[۱] (۱۹۸۷) شباهت‌های قابل توجه بین خانوادهای گوسی وارون و خانواده گوسی اشاره کرده‌اند. یورگنسن[۲](۱۹۹۰) تابع توزیع گوسی وارون تعمیم یافته را معرفی‌نمود وخواص آن را بیان‌کرد. مطالعه ای گسترده درباره پیدایش و خواص نظری توزیع IG توسط سشادری[۳] (۱۹۹۹) صورت گرفت. مادولکاروناتارجان[۴] (۲۰۰۲) خصوصیات این توزیع از جمله تقارن، چولگی وکشیدگی را بررسی کردند. آنها همچنین نتایج جذابی در مورد این توزیع و خصوصیات آن به‌دست‌آوردند. به عنوان مثال نشان دادند که توزیع گوسی وارون متعلق به خانواده نمایی است و درصورتی که داده‌ها نامنفی وچولگی مثبت باشند به عنوان جایگزینی برای توزیع نرمال استفاده می‌شود. توزیع گوسی وارون چند متغیره توسط مینامی[۵] (۲۰۰۴) معرفی شد او در سال ۲۰۰۷ نشان داد که توزیع لاگرانژ چند متغیره به توزیع گوسی وارون چند متغیره میل می‌کند. مادولکارو وانگ[۶] (۲۰۰۷) روابط بین Rمتقارن با توزیع گوسی وارون متقارن را بیان کردند. همچنین درباره توزیع $\frac{1}{\sqrt{X}}$ هنگامی که X دارای توزیع گوسی وارون است مطالعه کردند. سانهزا[۷]و همکاران (۲۰۰۸) قانون‌های بیضی شکل را که خانواده ای از مدل های احتمال متقارن رادر برمی‌گیرد، مورد مطالعه قرار دادند. این خانواده، توزیع های با دم های سبک وسنگین را در بر دارد. این مدل ها ممکن است برداده های حتی با دارای نقاط پرت به خوبی برازش داده شوند. آرنولد[۸]و سشادری (۲۰۰۹) خصوصیات دیگری از این توزیع را بیان‌کردند. بالاکیریشنان[۹]و همکاران (۲۰۰۹) به معرفی توزیع گوسی وارون آمیخته و تبدیل‌ها و گشتاورهای آن پرداختند و یک روش جدید برای محاسبه گشتاورهای توزیع معرفی کردند و از گشتاورهای توزیع برای انتخاب مدل مناسب استفاده نمودند.

یکی از کاربردهای مهم این توزیع در مسئله تحلیل داده‌های چوله است. در تحلیل داده‌ها، در بسیاری از موارد داده ها متقارن نبوده و دارای چولگی هستند. ولی با این وجود برای تحلیل آن‌ها از توزیع نرمال استفاده شود. اگرچه توزیع‌های چوله‌ای چون گاما، لگ‌نرمال و وایبل دامنه کاربردی وسیع خود را دارند، ولی وسعت کاربردی هیچ‌کدام از آن‌ها به اندازه توزیع نرمال نیست. به‌وسیله توزیع نرمال است که می‌توان برای داده‌ها تحلیل واریانس انجام داد، فاصله اطمینان ساخت، معادله رگرسیون برازاند.

---

1 Folks & Chhikara

2 Jorgensen,

3 Seshadri

4 Mudholkar & Natarajan

5 Minami

6 Mudholkar & Wang

7 Sanhueza

8 Arnold

9 Balakrishnan

هنگام مواجه شدن با داده‌های نامتقارن معمولاً سعی براین‌است که با نوعی تبدیل، دادها را به شکل متقارن درآورند. برای مثال از تبدیل کاکس-باکس برای رفع چولگی دادها استفاده می‌شود. اگرچه این کار ممکن است درست باشد، مثلاً در طرح آزمایش‌ها وارون متغیر پاسخ بهتر از متغیر پاسخ عمل کند. اما این مشکل باقی می ماند که تحلیل انجام شده بر اساس داده های تبدیل یافته است و نتایج بر حسب متغیر جدید ناشی از تبدیل به‌دست می‌آید نه متغیر اصلی پاسخ، که مطلوب نیست. برای حل این مشکل بهتر است در تحلیل دادهای چوله از توزیع‌های چوله استفاده نماییم.

۱-۲. فرایند وینر[۱]

یکی از مهم‌ترین فرایندهای تصادفی در احتمال کاربردی، فرایند وینر است که به صورت توصیفی از حرکت براونی در فیزیک سرچشمه گرفته‌است. این پدیده که در سال ۱۸۲۸توسط رابرت براون[۲] گیاه شناس انگلیسی کشف شد، عبارتست از حرکت ذرات کوچکی که در مایع یا گاز غوطه‌ور هستند. اولین تعریف این فرایند تصادفی که حرکت براونی را نیز دربرمی‌گیرد توسط وینر[۳] درسال ۱۹۱۸ بیان شده است. تعریف فرایند وینر یا به طور ساده فرایند حرکت براونی به صورت زیر است:

تعریف ۱-۱ فرایند تصادفی $\{X(t); t \geq 0\}$ با خواص زیر را یک فرایند وینر با رانش[٤] ν وواریانس σ^2 گویند.

۱.فرایند $X(t)$ دارای نموهای مستقل باشد یعنی برای هر $t_1 < t_2 < t_3 < t_4$ متغیرهای $X(t_2) - X(t_1)$ و $X(t_4) - X(t_3)$ مستقل باشند.

۲. $X(t_2) - X(t_1)$ دارای توزیع نرمال با میانگین $\nu(t_2 - t_1)$ و واریانس $\sigma^2(t_2 - t_1)$ باشد.

۱-۳. توزیع گوسی وارون

اگر متغیرتصادفی پیوسته و مثبت X دارای تابع چگالی احتمال زیر باشد

$$f(x;\mu,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right), \qquad x>0, \lambda>0, \mu>0 \qquad (۱-۱)$$

1 Wiener Process
2 Robert Brown
3 Wiener
4 Drift

آنگاه X دارای توزیع گوسی وارون است که منسوب به تویدی[1] می باشد .پارامتر μ میانگین توزیع و λ پارامتر مقیاس است.

تویدی با تعریف η و α به صورت $\eta = \dfrac{\lambda}{\mu}$ و $\alpha = \dfrac{1}{2\mu^2}$، سه شکل دیگر برای چگالی $f(x;\mu,\lambda)$ به‌صورت زیر عنوان نمود

$$f_1(x;\alpha,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left\{-\alpha\lambda x + \lambda(2\alpha)^{\frac{1}{2}} - \frac{\lambda}{2x}\right\}, \qquad x>0, \lambda>0, \alpha>0 \tag{۲-۱}$$

$$f_2(x;\mu,\eta) = \left(\frac{\mu\eta}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\eta x}{2x} + \eta - \frac{\mu\eta}{2x}\right), \qquad x>0, \mu>0, \eta>0 \tag{۳-۱}$$

$$f_3(x;\eta,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\eta^2 x}{2\lambda} + \eta - \frac{\lambda}{2x}\right), \qquad x>0, \lambda>0, \eta>0 \tag{۴-۱}$$

با قرار دادن $\mu = 1$ در چگالی (۱-۱) شکل استاندارد توزیع حاصل می‌شود که در واقع شکل حدی توزیع حجم نمونه در یک آزمون دنباله ای نسبت احتمال می‌باشد که توسط والد[2] به‌دست آمده است . از این رو گاهی اوقات این توزیع به عنوان توزیع والد استاندارد خوانده می‌شود. رابطه بین صورت مختلف تابع چگالی به‌صورت زیر است:

$$f(x;\mu,\lambda) = \mu^{-1} f_2(\frac{x}{\mu};1,\eta) = \lambda^{-1} f_3(\frac{x}{\lambda},\eta,1) \tag{۵-۱}$$

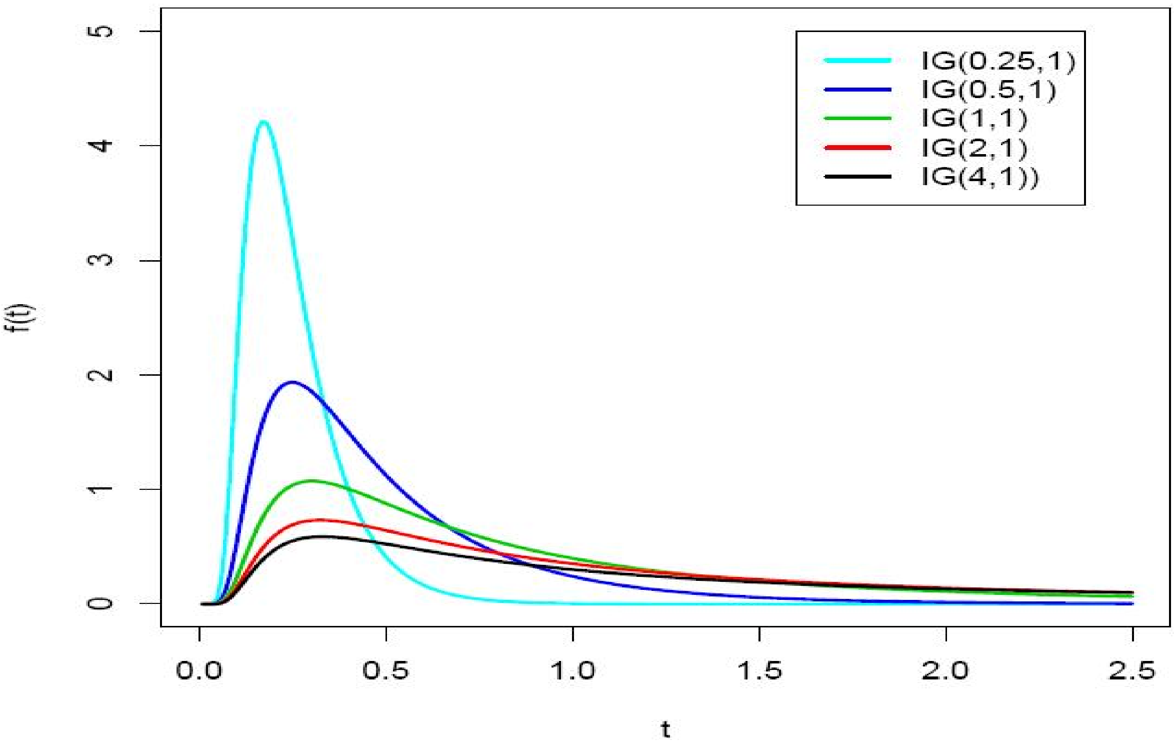
ضرایب چولگی و کشیدگی به ترتیب برابر $\dfrac{15}{\eta}$ و $\dfrac{3}{\sqrt{\eta}}$ هستند که فقط به پارامتر η بستگی دارند . بدین لحاظ به η پارامتر شکل توزیع نیز می گویند . برای مقادیر کوچک η چولگی تابع زیاد است و با افزایش η توزیع گوسی وارون به سمت نرمال میل می کند.تابع چگالی احتمال گوسی وارون تک مدی و چولگی آن مثبت است و بنابراین منحنی چگالی آن کشیده به راست می باشد . اشکال ۱-۱و ۱-۲ منحنی های نمایش تابع چگالی احتمال توزیع گوسی وارون برای $\mu = 1$ ومقادیر مختلف λ و برای $\lambda = 1$ و مقادیر مختلف μ را نشان می دهد . می توان نشان داد که مد یا نمای توزیع گوسی وارون به صورت زیر است :

$$x_{\text{mod}} = \mu\left[1 + \left(1 + \frac{9}{4\eta^2}\right)^{\frac{1}{2}} - \frac{3}{2\eta}\right] \tag{۶-۱}$$

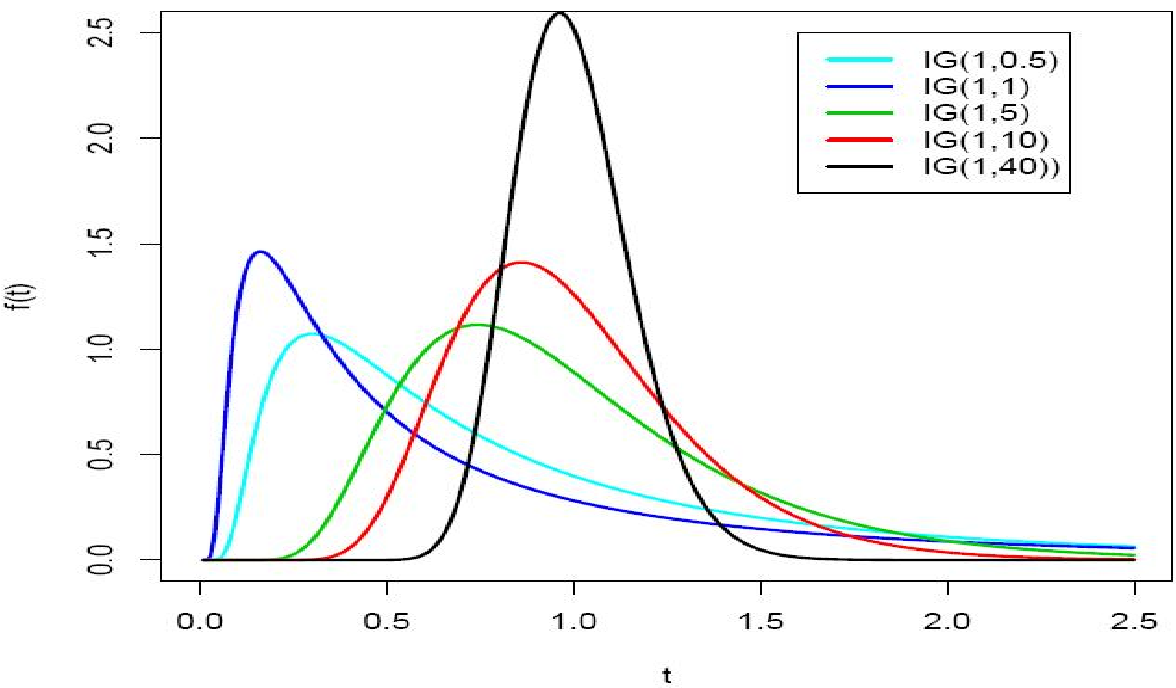
---

[1] Tweedie
[2] Wald

در این پایان نامه، از این به بعد چگالی ۱-۱ به عنوان شکل متداول توزیع گوسی وارون با پارامترهای μو λ و با نماد $IG(\mu,\lambda)$ مورد استفاده قرار می‌گیرد.



شکل(۱-۱) منحنی چگالی احتمال توزیع گوسی وارون برای $\lambda=1$ و۵ مقدار μ



شکل(۱-۲) منحنی چگالی احتمال توزیع گوسی وارون برای $\mu=1$ و۵ مقدار λ

۱-٤. تابع مولد گشتاور و تابع مشخصه توزیع گوسی وارون

اگر متغیر تصادفی X دارای توزیع گوسی وارون با پارامتر های μو λ باشد، در این صورت

$$M_X(t)=\int_0^\infty e^{tx}f(x;\mu,\lambda)\,dx=e^{\lambda/\mu}\int_0^\infty\left\{\frac{\lambda}{2\pi x^3}\right\}^{\frac{1}{2}}\exp\left\{-\frac{\lambda(x^2+\mu^2)-2\mu^2x^2t}{2\mu^2x}\right\}dx$$

حال می‌توان نوشت

$$-\frac{\lambda(x^2+\mu^2)-2\mu^2x^2t}{2\mu^2x}=\frac{-\lambda\left(\sqrt{1-\frac{2\mu^2t}{\lambda}}\,x-\mu\right)^2}{2\mu^2x}-\frac{\lambda}{\mu}\sqrt{1-\frac{2\mu^2t}{\lambda}}=\frac{-\lambda(x-\mu')^2}{2\mu'^2x}-\frac{\lambda}{\mu}\sqrt{1-\frac{2\mu^2t}{\lambda}}$$

که در آن $\mu'=\dfrac{\mu}{\left(1-\frac{2\mu^2t}{\lambda}\right)^{\frac{1}{2}}}$ است. با جایگذاری رابطه اخیر در انتگرال نمایش $M_X(t)$ داریم :

$$M_X(t)=\exp\left\{\frac{\lambda}{\mu}\right\}\exp\left\{\frac{-\lambda}{\mu}\sqrt{1-\frac{2\mu^2t}{\lambda}}\right\}\int_0^\infty\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left\{\frac{-\lambda(x-\mu')}{2\mu'^2x}\right\}dx$$

تابع زیر انتگرال بالا تابع چگالی احتمال یک توزیع گوسی وارون با پارامترهای μ'و λ است. بنابراین مقدار انتگرال برابر یک و در نتیجه تابع مولد گشتاورتوزیع گوسی وارون به صورت زیر می باشد

$$M_X(t)=\exp\left\{\frac{\lambda}{\mu}\left[1-\left(1-\frac{2\mu^2t}{\lambda}\right)^{\frac{1}{2}}\right]\right\},\qquad t<\frac{\lambda}{2\mu^2} \tag{۷-۱}$$

به طریق مشابه می توان نشان داد که تابع مشخصه یک توزیع گوسی وارون با پارا متر های μ و λ به صورت زیر می باشد.

$$C_X(t)=\exp\left\{\frac{\lambda}{\mu}\left[1-\left(1-\frac{2i\mu^2t}{\lambda}\right)^{\frac{1}{2}}\right]\right\},\qquad t\in R \tag{۸-۱}$$

با قرار دادن $t=0$ در مشتق مرتبه $r-$ ام تابع مولد گشتاور $M_X(t)$ گشتاور حول مبدأ مرتبه $r-$ ام طبق رابطه زیر قابل محاسبه است:

$$E(X^r)=\mu^r\sum_{s=0}^{r-1}\frac{(r-1+s)!}{s!(r-1-s)!}\left(\frac{2\lambda}{\mu}\right)^{-s} \tag{۹-۱}$$

پس به ازای $r=1$ و $r=2$ داریم :

$$E(X)=\mu \qquad E(X^2)=\frac{\mu^3}{\lambda}+\mu^2 \quad \Rightarrow \quad Var(X)=\frac{\mu^3}{\lambda}$$

با قرار دادن $\sigma=\frac{\mu^3}{\lambda}$، چگالی (۱-۱) را می توان بر حسب پارامترهای میانگین μ و σ انحراف معیار توزیع به صورت زیر باز نویسی کرد:

$$f(x;\mu,\lambda)=\frac{\mu}{\sigma}\left(\frac{\mu}{2\pi x^3}\right)^{\frac{1}{2}}\exp\left(\frac{-\mu(x-\mu)^2}{2\sigma^2 x}\right) \qquad , \; x>0,\sigma>0,\mu>0 \tag{۱۰-۱}$$

۵-۱. متغیر های تصادفی وارون و علت نام گذاری توزیع

اگر $M_X(t)$ نشان دهنده تابع مولد گشتاور متغیر تصادفی X باشد. تابع مولد انباشتگی[1] متغیر تصادفی X به صورت زیر تعریف می‌شود:

$$\psi(t)=\log M_X(t)=\log E[e^{tX}] \tag{۱۱-۱}$$

حال اگر $L_X(t)$ نشان دهنده لگاریتم تبدیل لاپلاس متغیر تصادفی نامنفی X باشد . بنابراین می توان نوشت :

$$L_X(t)=\log E[e^{-tx}]=\psi(-t) \qquad t>0 \tag{۱۲-۱}$$

در سال ۱۹۴۵ تویدی ارتباطی وارون بین تابع مولد انباشتگی جفت‌هایی از متغیر های تصادفی دو جمله ای و دو جمله ای منفی، پواسن و گاما،و نرمال و گوسی وارون ارائه نمود.

در این‌جا ابتدا در حالت کلی متغیرهای تصادفی وارون را معرفی می‌شوند. سپس این ارتباط وارون در مورد جفت متغیر های تصادفی که دارای این خاصیت هستند مورد بررسی قرار گیرند.

تعریف ۱-۲ متغیرهای تصادفی X و Y وارون یکدیگرند اگر وتنها اگر $L_X(t)$ و $L_Y(t)$ به ازای جمیع مقادیر t در دامنه مشترک در شرایط زیر صدق کنند.

$$i)\ L_X(t)=\alpha L(t) \qquad\qquad ii) L_Y(t)=\beta L^{-1}(t) \tag{۱۳-۱}$$

که در آن α و β مقادیری ثابت اند و $L^{-1}(t)$ وارون تابع $L(t)$ می باشد یعنی :

$$L(L^{-1}(t))=t$$

توزیع های متناظر با متغیرهای تصادفی وارون را توزیع های وارون می گویند.

---

[1] Cumulant Generating Function

مثال ۱-۱: اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 با تابع چگالی احتمال زیر باشد :

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}\ ,\qquad x\in R,\mu>0,\sigma>0$$

آن گاه:

$$E[e^{tx}]=\exp\left\{\mu t+\frac{\sigma^2t^2}{2}\right\}\ \Rightarrow\ L_X(t)=\mu\left(\frac{\sigma^2t^2}{2\mu}-t\right)\ ,\ t>0$$

با فرض $\mu\ =\ \alpha$ و $L(t)=\dfrac{\sigma^2t^2}{2\mu}-t$ به راحتی می توان ملاحظه نمود که :

$$L^{-1}(t)=\frac{\mu}{\sigma^2}\left[1-\left(1+\frac{2\sigma^2}{\mu}t\right)^{\frac{1}{2}}\right]\ ,\qquad t<\frac{\mu}{2\sigma^2}$$

با قرار دادن $\lambda=\dfrac{\mu^3}{\sigma^2}$ داریم :

$$L^{-1}(t)=\frac{\lambda}{\mu^2}\left[1-\left(1+\frac{2\mu^2}{\lambda}t\right)^{\frac{1}{2}}\right]\ ,\qquad t<\frac{\lambda}{2\mu^2}$$

حال اگر متغیر تصادفی Y دارای تابع چگالی احتمال ۱-۱ یعنی دارای توزیع گوسی وارون باشد .
آن گاه:

$$L_Y(t)=\frac{\lambda}{\mu}\left[1-\left(1+\frac{2\mu^2}{\lambda}t\right)^{\frac{1}{2}}\right]=\mu L^{-1}(t)\ ,\qquad t<\frac{\lambda}{2\mu^2}$$

با فرض $\mu\ =\ \beta$ به ازای جمع مقادیر $t<\dfrac{\lambda}{2\mu^2}$ روابط **(۱-۹)** در مورد متغیر های تصادفی X و Y حفظ می شود یعنی متغیر تصادفی Y وارون متغیر تصادفی نرمال (گوسی) است.

با وجود چنین ارتباطی بین توابع $L_X(t)$ و $L_Y(t)$، تویدی نام گوسی وارون را برای متغیر تصادفی Y پیشنهاد نمود.