بسم الله الرحمن الرحيم

دانشکده ریاضی ورایانه

بخش آمار

پایان نامه تحصیلی برای دریافت درجه کارشناسی ارشد

رشته آمار ریاضی

برآوردپارامترهای توزیع نمایی تعمیم یافته

استاد راهنما:

دکتر احدجمالیزاده

استاد مشاور:

دکتر علیرضاعرب پور

مؤلف:

حمید رضا ابراهیمی

بهمن ماه ۱۳۸۹

این پایان نامه به عنوان یکی از شرایط درجه کارشناسی ارشد به

گروه آمار

دانشکده ریاضی و کامپیوتر

دانشگاه شهید باهنر کرمان

تسلیم شده است و هیچگونه مدرکی به عنوان فراغت از تحصیل دوره مزبور شناخته نمی شود.

دانشجو : حمید رضا ابراهیمی

استاد راهنما : دکتر احد جمالی زاده

استاد مشاور : دکتر علیرضا عرب پور

داور ۱ : دکتر وحید امیرزاده

داور۲ : دکترمحمد علی یعقوبی

معاونت پژوهشی و تحصیلات تکمیلی دانشکده : دکتر سید ناصر حسینی

اگر شایسته تقدیم باشد، به

خانواده، اساتید و دوستانم،

که مشوق من درامر تحصیل بوده اند.

قدردانی :

اول سپاس بیکران خدایی راکه هرتوفیقی درگروعنایت اوست و سپس اساتید بزرگواری که مرا درس زندگی آموختند.

بهمن ماه ۱۳۸۹

چکیده :

توزیع نمایی تعمیم یافته نسبت به توزیع های گاما، وایبل و لوگ نرمال انعطاف پذیری بیشتری دارد بنابراین درآنالیز داده های طول عمرمناسبتر است.

در این پایان نامه توزیع نمایی تعمیم یافته و ویژگی های آن را ارائه می دهیم و روشهای مختلف برآورد پارامترهای توزیع نمایی تعمیم یافته نظیر برآوردگرهای درستنمایی ماکزیمم، برآوردگرهای گشتاوری، برآوردگرهای صدکی، برآوردگرهای حداقل مربعات و حداقل مربعات وزنی و همچنین برآوردگرهای L-گشتاوری را ارائه داده وآنها را براساس مقادیر حجم نمونه و پارامتر شکل مقایسه می کنیم.

ودر پایان توزیع نمایی تعمیم یافته دو متغیره را معرفی کرده و پارامترهای آن را به روش درستنمایی ماکزیمم برآورد می کنیم.

کلمات کلیدی :توزیع نمایی تعمیم یافته، توریع نمایی تعمیم یافته دومتغیره ، برآوردگرهای درستنمایی ماکزیمم ، برآوردگرهای L-گشتاوری، برآوردگرهای حداقل مربعات.

فهرست مطالب

عنوان **صفحه**

فصل دوم- روشهای برآوردپارامترهای توزیع نمایی تعمیم یافته و مقایسه آنها

فصل سوم- توزیع نمایی تعمیم یافته دو متغیره $(BVGE)$

فصل اول

توزیع نمایی تعمیم یافته

۱-۱ مقدمه

در قرن نوزدهم توابع توزیع تجمعی خاص توسط گومپرتز و ورهالست[۱] [۳] برای مقایسه جداول مرگ و میر انسانی و رشد مرگ و میر به کار می‌رفت که یکی ازآنها به صورت زیراست:

$$G(t)=(1-\rho e^{-t\lambda})^{\alpha}, \quad t>\frac{1}{\lambda} ln\rho . \qquad (۱,۱,۱)$$

به طوری که ρ و λو α اعداد حقیقی مثبت و معلوم هستند

در این فصل توزیع نمایی تعمیم یافته دو پارامتری را معرفی می کنیم و خواص این توزیع را بیان می کنیم .

همچنین چولگی و برجستگی توزیع نمایی تعمیم یافته دو پارامتری و تابع مخاطره و آماره های ترتیبی آن را بیان می کنیم.

توزیع نمایی تعمیم یافته دو پارامتری

توزیع نمایی تعمیم یافته دو پارامتری به صورت زیر تعریف می شود [٦]

تعریف ۱-۲ متغیر تصادفی X دارای توزیع نمایی تعمیم یافته دو پارامتری است اگر تابع توزیع آن به صورت زیر باشد:

$$F(x;\alpha,\lambda)=(1-e^{-\lambda x})^{\alpha}, \ x>0, \ \alpha>0, \lambda>0 . \qquad (۲,۱,۱)$$

به طوری که α و λ به ترتیب نقش پارامتر شکل و مقیاس را بازی می‌کنند.

اگر X یک متغیر تصادفی با تابع توزیع نمایی تعمیم یافته $F(x;\alpha,\lambda)=(1-e^{-\lambda x})^{\alpha}, \quad x>0$ باشد. آنگاه، تابع چگالی آن به صورت زیر بدست می‌آید:

$$f(x;\alpha,\lambda)=\alpha\lambda(1-e^{-\lambda x})^{\alpha-1}e^{-\lambda x}, \qquad x>0, \ \alpha>0, \ \lambda>0.$$

[۱] - Gompertz and Verhulst

که با نماد $GE(\alpha,\lambda)$ [1]نشان داده می‌شود.توابع چگالی نمایی تعمیم یافته به ازای مقادیر مختلف α دارای شکل‌های مختلفی است.

توزیع نمایی تعمیم یافته دو پارامتری یک حالت خاص از توزیع وایبل گسترش یافته سه پارامتری مطرح شده توسط مودولکار و همکاران[2] [۲۳] می‌باشد که در آنالیز داده‌های طول عمر به خصوص به جای توزیع‌های وایبل دو پارامتری و گامای دو پارامتری کاملاً مؤثر بکار می‌رود.اگر $\alpha = 1$باشد بر توزیع نمایی یک پارامتری منطبق‌می‌شود.بنابراین توزیع‌های نمایی تعمیم یافته، وایبل و گاما تعمیم‌هایی از توزیع نمایی یک پارامتری می‌باشند.

توزیع نمایی تعمیم یافته دارای ویژگی‌های فیزیکی مطلوبی می‌باشد. یک سیستم موازی تشکیل شده از n عضورا در نظر بگیریدکه سیستم زمانی کار می‌کند که حداقل یکی از عضوها کار کند. اگر توزیع طول عمر عضوها، متغیرهای تصادفی نمایی از هم مستقل و هم توزیع (iid)باشند. آنگاه توزیع طول عمر سیستم به صورت زیر بدست می‌آید:

$$F(x;n,\lambda) = (1-e^{-\lambda x})^n, \quad x > 0, \ \lambda > 0.$$

ملاحظه می‌کنیم که رابطه مذکور تابع توزیع نمایی تعمیم یافته با $\alpha = n$ را نشان می‌دهد. بنابراین برخلاف تابع توزیع وایبل که یک سیستم دنباله‌ای را نشان می‌دهد، تابع توزیع نمایی تعمیم یافته یک سیستم موازی را نشان می‌دهد.

به دلیل شکل مناسب تابع توزیع، متغیر تصادفی نمایی تعمیم یافته را به آسانی می‌توان تولید کرد. به عنوان مثال اگر U نشان دهنده یک متغیر تصادفی یکنواخت از $[0,1]$ باشد. آنگاه $X = -(1/\lambda)\, ln(1-U^{1/\alpha})$ دارای توزیع نمایی تعمیم یافته با تابع توزیع $F(x;\alpha,\lambda) = (1-e^{-\lambda x})^{\alpha}, \quad \alpha > 0, \ \lambda > 0$ می‌باشد.

یکی دیگر از مزیت‌های توزیع نمایی تعمیم یافته این است که در آنالیز داده‌های طول عمر چوله، کارایی زیادی دارد.

[1] Generalized Exponential Distribution
[2] - Mudholkar and Srivastava

توزیع نمایی تعمیم یافته یک حالت خاص از رابطه (۱,۱,۱) می باشد به طوری که $\rho = 1$ درنظر گرفته شده است.

۱- ۲-۱ خواص توزیع نمایی تعمیم یافته

دراین بخش خواص توزیع نمایی تعمیم یافته را بیان می کنیم:

۱- اگر $\alpha \leq 1$ تابعی نزولی و به ازای $\alpha > 1$ تک مدی، چوله به راست، مشابه توابع چگالی وایبل یا گاما می‌باشد و حتی به ازای پارامتر شکل خیلی بزرگ هم متقارن نیست.

۲- به ازای $\lambda = 1$ مد به سمت $log(\alpha)$ میل می‌کند.وبه ازای $\alpha \leq 1, \alpha > 1$ مد نزدیک صفر می‌باشد و میانه آن برابر $-ln(1-(.5)^{1/\alpha})$ است.

۳- میانگین، میانه و مد توابعی غیر خطی از پارامتر شکل هستند و چنانچه پارامتر شکل به سمت بی‌نهایت میل کند، همه اینها به سمت بی‌نهایت میل می‌کنند.

۴- برای مقادیر بزرگ α، میانگین، میانه و مد تقریباً برابر با $log(\alpha)$ می‌باشند.

گشتاورهای مختلف توزیع نمایی تعمیم یافته، با استفاده از تابع مولد گشتاورآن بدست می آیند

۵- اگر X دارای توزیع $GE(\alpha, \lambda)$ باشد، آنگاه تابع مولد گشتاور $M(t)$ برای $t < \lambda$ به صورت زیر بدست می‌آید:

$$M(t) = E(e^{tX}) = \frac{\Gamma(\alpha+1)\Gamma(1-t/\lambda)}{\Gamma(\alpha-t/\lambda+1)}$$

بنابراین با استفاده از رابطه بالا داریم:

$$E(X) = \frac{1}{\lambda}\left[\psi(\alpha+1) - \psi(1)\right],$$

$$V(X) = \frac{1}{\lambda^2}\left[\psi'(1) - \psi'(\alpha+1)\right].$$

به طوری‌که $\psi(.)$ و مشتقات آن توابع گاما دوگانه و گاما چندگانه می‌باشند. که از رابطه زیر به دست می آیند.

$$(d/dt)\Gamma(1-t) = -\psi(1-t)\Gamma(1-t)$$

$$\Rightarrow \psi(x) = \left((d(\log\Gamma(x)))/dx\right)$$

۶-اگر λ ثابت و α صعودی باشند. میانگین توزیع نمایی تعمیم یافته به سمت ∞ میل می‌کند.

۷- به ازای λ ثابت، واریانس نیز صعودی است و به سمت $\pi^2/6\lambda$ میل می‌کند، این ویژگی در مقایسه با توزیع گاما و وایبل کاملاً متفاوت است. در مورد توزیع گاما چنانکه پارامتر شکل صعودی باشد، واریانس به سمت ∞ میل می‌کند.در حالی‌که در مورد توزیع وایبل، واریانس به ازای مقادیر بزرگ پارامتر شکل α تقریباً برابر با $\pi^2/6\lambda\alpha^2$ است. برای جزئیات بیشتربه گوپتا و کاندو[1] [۵] رجوع شود.

یک نمایش تصادفی از $GE(\alpha,1)$ برای محاسبه گشتاورهای مختلف توزیع نمایی تعمیم یافته مورد استفاده قرار می گیرد.

۸- اگر α یک عدد صحیح مثبت مانند n باشد، آنگاه توزیع X باتوزیع $\sum_{j=1}^{n} Y_j/j$ یکی می‌باشد. به طوریکه Y_j ها متغیرهای تصادفی نمایی iid با میانگین یک می‌باشند.

۹- اگر α یک عدد صحیح نباشد، آنگاه توزیع X با توزیع $\sum_{j=1}^{[\alpha]} \frac{Y_j}{j+\langle\alpha\rangle} + Z$. یکی می‌باشد، به طوری که $\langle\alpha\rangle$ جزء کسری و $[\alpha]$ جزء صحیح α را نشان می‌دهند و Z دارای توزیع $GE(\langle\alpha\rangle,1)$ می‌باشد که از Y_j ها مستقل است. [۶]

۱-۲-۲چولگی و برجستگی توزیع نمایی تعمیم یافته

چولگی و برجستگی توزیع نمایی تعمیم یافته به صورت زیر بدست می‌آیند:

$$\text{چولگی} = \sqrt{\gamma_1} = \frac{\mu_3}{\mu_2^{3/2}} \quad ,$$

$$\text{برجستگی} = \gamma_2 = \frac{\mu_4}{\mu_2^2} .$$

[1] Gupta and Kundu

به طوری‌که μ_4, μ_3, μ_2 به ترتیب، دومین، سومین و چهارمین گشتاورها می‌باشند که به صورت جملاتی از توابع گاما دوگانه و گاما چند گانه به صورت زیر نوشته می شوند:

$$\mu_2 = \frac{1}{\lambda^2}\left[\psi(1) - \psi'(\alpha+1) + \left(\psi(\alpha+1) - \psi(1)\right)^2\right]$$

$$\mu_3 = \frac{1}{\lambda^3}\left[\psi''(\alpha+1) - \psi''(1) + 3\left(\psi(\alpha+1) - \psi(1)\right)\left(\psi'(1) - \psi'(\alpha+1)\right) + \left(\psi(\alpha+1) - \psi(1)\right)^3\right]$$

$$\mu_4 = \frac{1}{\lambda^4}\Big[\psi'''(1) - \psi'''(\alpha+1) + 3\left(\psi'(1) - \psi'(\alpha+1)\right)^2 + 4\left(\psi(\alpha+1) - \psi(1)\right)\left(\psi''(\alpha+1) - \psi''(1)\right) +$$

$$6\left(\psi(\alpha+1) - \psi(1)\right)^2\left(\psi'(1) - \psi'(\alpha+1)\right) + \left(\psi'''(1) - \psi'''(\alpha+1)\right)^4\Big]$$

به صورت عددی نشان داده شده است که چولگی و برجستگی هر دو توابعی نزولی از α هستند. علاوه بر این حد مقدار چولگی تقریباً برابر ۱/۱۳۹۵۴۷ می‌باشد.

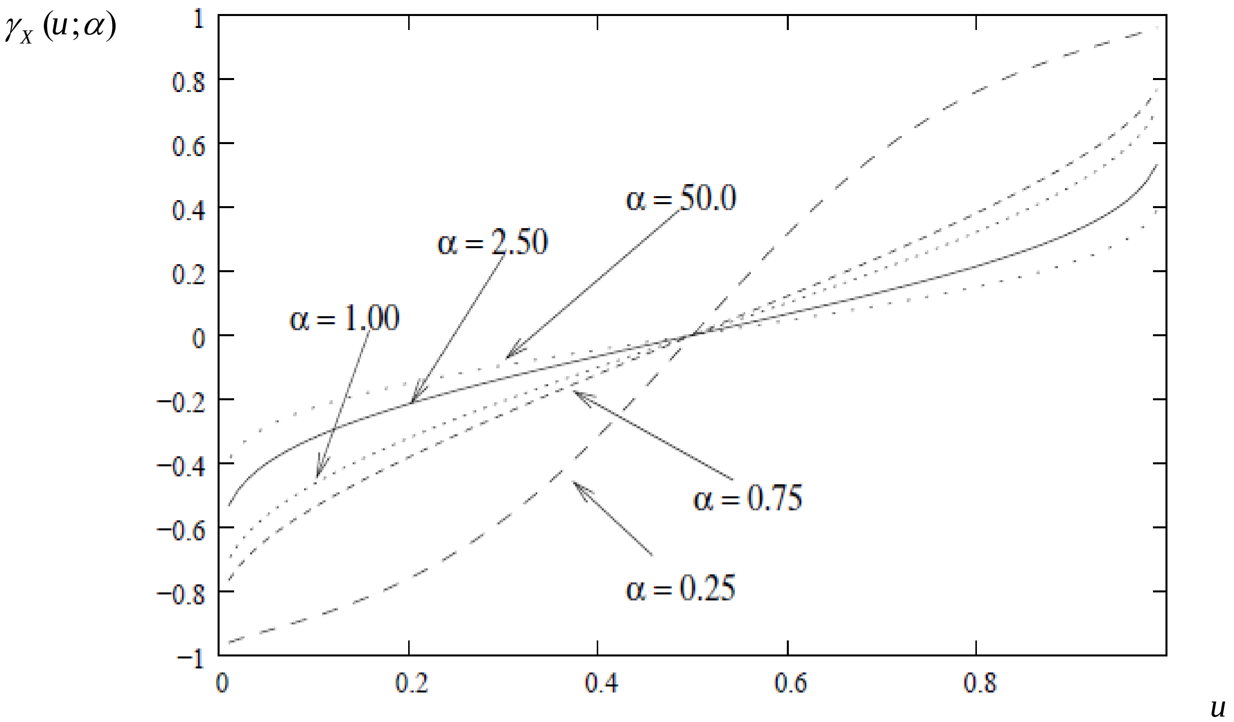
توزیع نمایی تعمیم یافته دارای تابع چولگی مک گیلواری[1] [۱۵] به شکل زیر می‌باشد:

$$\gamma_X(u;\alpha) = \frac{\ln(1-u^{1/\alpha}) + \ln(1-(1-u^{1/\alpha})) - 2\ln(1-(1/2)^{1/\alpha})}{ln(1-u^{1/\alpha}) - ln(1-(1-u^{1/\alpha}))}. \qquad (۱,۲,۱)$$

شکل توابع چولگی رابطه فوق از توزیع نمایی تعمیم یافته به ازای مقادیر مختلف α در شکل ۱-۱ آورده شده است.

---

[1] Mac Gillvary



شکل۱-۱ تابع چولگی مک گیلواری برای توزیع نمایی تعمیم یافته

با استفاده از شکل ۱-۱نتیجه می گیریم که چولگی به ازای مقادیر بزرگ α تغییر معناداری ندارد.

۱-۳ توابع مخاطره

تابع مخاطره توزیع نمایی تعمیم یافته به صورت زیر بدست می‌آید:

$$h(x;\alpha,\lambda)=\frac{f(x;\alpha,\lambda)}{1-F(x;\alpha,\lambda)}=\frac{\alpha\lambda e^{-\lambda x}(1-e^{-\lambda x})^{\alpha-1}}{1-(1-e^{-\lambda x})^{\alpha}}. \qquad (۱,۳,۱)$$

چون λ پارامتر مقیاس است، شکل تابع مخاطره‌ای به λ بستگی ندارد و فقط به α بستگی دارد. اگر λ ثابت فرض شود، توزیع نمایی تعمیم یافته به ازای $\alpha>1$ دارای تابع مخاطره صعودی و به ازای $\alpha<1$ تابع مخاطره نزولی می‌باشد و به ازای $\alpha=1$ تابع مخاطره توزیع نمایی تعمیم یافته ثابت است.

رفتار تابع مخاطره توزیع نمایی تعمیم یافته عیناً مانند تابع مخاطره توزیع گاما می‌باشد که با تابع مخاطره توزیع وایبل کاملاً متفاوت است.

عکس تابع مخاطره توزیع نمایی تعمیم یافته به صورت زیر بدست می‌آید:

$$\gamma(x;\alpha,\lambda)=\frac{f(x;\alpha,\lambda)}{F(x;\alpha,\lambda)}=\frac{\alpha\lambda e^{-\lambda x}}{1-e^{-\lambda x}}. \qquad (۲,۳,۱)$$

به ازای همه مقادیر α عکس تابع مخاطره، تابعی نزولی از x می‌باشد.

می‌دانیم که عکس تابع مخاطره توزیع نمایی به صورت $\frac{\lambda e^{-\lambda x}}{1-e^{-\lambda x}}$ می‌باشد. بنابراین عکس تابع مخاطره توزیع نمایی تعمیم یافته با عکس تابع مخاطره توزیع نمایی متناسب است.

۱-۴ آماره های ترتیبی

فرض کنید $X_1,...,X_n$ متغیرهای تصادفی نمایی تعمیم یافته (iid) با پارامتر شکل α و پارامتر مقیاس $\lambda = 1$ باشند. همچنین $X_{(1)}<...<X_{(n)}$ آمارهای ترتیبی از این n متغیر تصادفی باشند. آنگاه تابع چگالی بزرگ ترین آماره ترتیبی $X_{(n)}$ به صورت زیر بدست می‌آید:

$$f_{X_{(n)}}(x;\alpha)=n\alpha e^{-x}(1-e^{-x})^{n\alpha-1}. \qquad (۱,۴,۱)$$

بنابراین $X_{(n)}$ هم دارای توزیع نمایی تعمیم یافته با پارامتر شکل $n\alpha$ و پارامتر مقیاس یک می‌باشد.

آماره‌های ترتیبی مختلف توزیع نمایی تعمیم یافته توسط راکب و احسن‌اله[۱][۱۶] مورد بررسی قرار گرفته است.توابع مولد گشتاور آماره‌های ترتیبی مختلف و گشتاورهای بدست آمده برای محاسبه بهترین برآوردگرهای خطی نااریب پارامترهای مکان و مقیاس توزیع نمایی تعمیم یافته به کارمی روند.

---

[۱] - Raqab and Ahsanullah

۱-۵ توزیع حاصل جمع

ازآنجا که تابع مولد گشتاور توزیع نمایی تعمیم یافته شکل مناسبی ندارد، بنابراین توزیع مجموع n متغیر تصادفی نمایی تعمیم یافته هم توزیع و از هم مستقل به سادگی بدست نمی‌آید.

همچنین می‌دانیم که اگر X دارای توزیع $GE\,(\alpha,1)$ باشد، آنگاه e^{-x} دارای توزیع بتا می‌باشد.

گوپتا و کاندو نشان داده اند که توزیع مجموع n متغیر تصادفی نمایی تعمیم یافته به صورت ترکیب نامتناهی از توزیع‌های نمایی تعمیم یافته نوشته می شود. در واقع به صورت آمیخته‌ای از ضرایب و پارامترهای متناظر توزیع نمایی تعمیم یافته بدست می آید.

۱-۶ نزدیکی یا شباهت با دیگر توزیع‌ها

توزیع نمایی تعمیم یافته درحالت کلی در تناوب با توزیع‌های گاما و وایبل معرفی شده است. در واقع می‌توان گفت در بیشتر مواقع توزیع نمایی تعمیم یافته در آنالیز داده‌های مثبت به جای توزیع‌های گاما، وایبل یا لوگ نرمال کاملاً مؤثر است.

گوپتا و کاندو شباهت‌های توزیع نمایی تعمیم یافته با توزیع‌های وایبل، گاما و لوگ نرمال را بررسی کرده اند. مشاهده شده است که برای حوزه خاصی از پارامترهای شکل، فاصله بین توزیع نمایی تعمیم یافته با توزیع‌های وایبل، گاما و لوگ نرمال خیلی جزئی است.

در اینجا دو سؤال مهم مطرح می‌شود:

۱) برای یک مجموعه داده معین، کدام توزیع را ترجیح می‌دهید؟

۲) می‌نیمم حجم نمونه مورد نیاز برای تشخیص بین دو تابع توزیع مناسب چه اندازه‌ای است؟

مسئله اول یک مسئله کلاسیک در آنالیز داده‌های آماری می‌باشد.

برای اولین بار کاکس[1] [۱] ، آزمون نسبت درستنمایی را برای تشخیص تابع توزیع بهتر پیشنهاد کرد و بخش عمده‌ای از تحقیق او در این زمینه اختصاص یافته است.

متأسفانه بنا نهادن یک نظریه جامع کار مشکلی است. بنابراین برای هر دو تابع توزیع معین نیازمند آزمون خاصی هستیم. مشاهده شده است که آماره آزمون نسبت درستنمایی جهت آزمون بین توزیع GE و سایر توزیع‌ها به طور مجانبی دارای توزیع نرمال است.جزئیات بیشتر در گوپتا و کاندو [۵ و۱۰] موجوداست.

۱-۶-۱ اندازه نمونه

اکنون در مورد سؤال دوم، به می‌نیمم حجم نمونه لازم برای تشخیص بین دو تابع توزیع مناسب می‌پردازیم.

این سؤال اهمیت زیادی دارد، هرچند که دو تابع توزیع مناسب به طور مجانبی همیشه قابل تشخیص هستند اما برای نمونه متناهی ممکن است تشخیص بین دو تابع توزیع سخت باشد.

به طور شهودی واضح است که اگردو تابع توزیع خیلی به هم نزدیک باشند، برای اینکه بتوانیم آنها را از یکدیگرتشخیص دهیم، به حجم نمونه خیلی بزرگی نیازمندیم. و اگر دو تابع توزیع کاملاً متفاوت باشند، آن گاه ممکن است نیازی به حجم نمونه بزرگ برای تشخیص بین این دو نداشته باشیم.علاوه براین اگردو تابع توزیع خیلی بهم نزدیک باشند، از نظر کاربردی ممکن است نیازی به فرق گذاشتن بین آنها نداشته باشیم.

بنابراین انتظار می‌رود که کاربر حد تحمل فاصله بین دو تابع توزیع را معین کند.حد تحمل به سادگی نشان می‌دهد که کاربر نباید بین دو تابع توزیع،فرق بگذارد اگر فاصله آنها کمتر از حد تحمل باشد.

برای کاربر، حد تحمل می‌نیمم حجم نمونه مورد نیاز برای تشخیص بین توزیع نمایی تعمیم یافته و سایر توزیع‌ها، توسط گوپتا و کاندو [۵ و ۱۰] معین شده است.

---

[1] - Cox

۱-۶-۲ تولید داده

از آنجا که توابع توزیع نمایی تعمیم یافته و گاما یا لوگ نرمال خیلی به هم نزدیک هستند، این ویژگی برای تولید متغیرهای تصادفی نرمال یا گاما با استفاده از توزیع نمایی تعمیم یافته بکار می‌رود. توابع توزیع گاما یا نرمال توابع معکوس ساده‌ای ندارند. بنابراین برای تولید اعداد تصادفی نرمال یا گاما به الگوریتم خاصی نیازمندیم.

در مورد توزیع گاما، ملاحظه شده است که اگر پارامتر شکل کمتر از ۲/۵ باشد، آنگاه می‌تواند با استفاده از توزیع نمایی تعمیم یافته تولید شود.

مقادیر دقیق متناظر پارامترهای شکل و مقیاس توزیع نمایی تعمیم یافته در گوپتا و کاندو[۱۱] موجود است.

توسط کاندو و گوپتا [۱۷] ملاحظه شده که توابع توزیع GE(۱۲/ ۹, ۱) و تابع توزیع لوگ نرمال با پارامترهای شکل و مقیاس به ترتیب ۰/۳۸۰۷۴۸۲ و ۲/۹۵۰۸۶۷۲ غیر قابل تشخیص هستند.

بنابراین توزیع لوگ نرمال با پارامترهای شکل و مقیاس متناظر می‌تواند با استفاده از توزیع نمایی تعمیم یافته تولید شود.گوپتا و همکاران اعداد تصادفی نرمال را با استفاده از توزیع نمایی تعمیم یافته تولید کرده‌اند. تعمیم شبیه‌سازی ها بیان می‌دارد که این شیوه ساده برای تولید اعداد تصادفی نرمال کاملاً مؤثر است.

همچنین می‌توانیم تابع توزیع نرمال استاندارد را با استفاده از تابع توزیع نمایی تعمیم یافته خیلی خوب به صورت زیر تقریب بزنیم. (تا سه رقم اعشار)

$$\Phi(z) \approx (1-e^{-e^{1.0792510+0.3820198z}})^{12.8}$$