مجله علوم آماری، پاییز و زمستان ۱۳۹۵

جلد ۱۰، شماره ۲، ص ۳۲۹-۳۴۴

DOI: 10.18869/acadpub.jss.10.2.329

اثبات رابطه سرگئی وینزکی برای تقریب تابع توزیع نرمال استاندارد

شهرام منصوری

دانشکده علوم ریاضی، دانشگاه شهید بهشتی

تاریخ دریافت: ۱۳۹۴/۵/۲۳ تاریخ آخرین بازنگری: ۱۳۹۵/۲/۲۲

چکیده : در بین تمام توزیع‌های آماری توزیع نرمال استاندارد مهمترین و کاربردی‌ترین توزیع آماری بوده و محاسبه سطح زیر منحنی چگالی و تابع توزیع آن مورد نیاز است. ضابطه این تابع بصورت یک انتگرال معین بیان می‌شود، ولی متاسفانه تابع اولیه آن دارای شکل بسته و تحلیلی نیست، لذا باید آن را تقریب زد. در این مقاله رابطه تقریبی سرگئی وینزکی با یک روش جدید اثبات می‌شود، سپس این تقریب با تغییراتی در رابطهٔ آن بهبود داده و نشان می‌دهیم حداکثر مقدار خطای آن کمتر از ۰/۰۰۰۰۵۸۴ است. در انتها رابطه‌ای نیز برای محاسبه صدک‌های توزیع نرمال به‌دست آورده می‌شود.

واژه‌های کلیدی : تابع توزیع نرمال استاندارد تابع خطا، تقریب

آدرس الکترونیک مسئول مقاله: شهرام منصوری، sh_mansouri@sbu.ac.ir

کد موضوع‌بندی ریاضی (۲۰۱۰): ۶۲E۱۷، ۶۲E۹۹

۱ مقدمه

تابع خطا و تابع توزیع احتمال نرمال استاندارد به‌ترتیب به‌صورت

$$\begin{aligned} erf(x) &= \frac{۲}{\sqrt{\pi}}\int_{۰}^{x} e^{-t^{۲}}dt; \qquad -\infty < x < \infty \\ \Phi(z) &= \int_{-\infty}^{z}\frac{۱}{\sqrt{۲\pi}}e^{-\frac{۱}{۲}t^{۲}}dt; \quad -\infty < z < \infty \end{aligned} \tag{۱}$$

تعریف شده‌اند، که در احتمال، آمار و معادلات دیفرانسیل با مشتقات جزئی کاربرد دارند. به‌عنوان مثال معادله نفوذ $\frac{\partial U}{\partial t} = a\frac{\partial^{۲}U}{\partial x^{۲}} - b\frac{\partial U}{\partial x}$ در قلمرو $t > ۰,\ x > ۰$ یا $t > ۰,\ x \in R$ به ازای شرایط مرزی خاص بر حسب تابع خطا بیان می‌شود. به‌راحتی می‌توان ثابت کرد که بین این دو تابع روابط زیر برقرارند.

$$\begin{aligned} \Phi(z) &= \frac{۱}{۲}(۱ + erf(\frac{z}{\sqrt{۲}})) \\ erf(x) &= ۲\Phi(\sqrt{۲}x) - ۱ \end{aligned} \tag{۲}$$

برخی ویژگی های تابع خطا عبارتند از

الف - $erf(x)$ تابعی فرد است.

ب - $|erf(x)| < ۱$

ج - $erf(۰) = ۰$

د - $erf(\infty) = ۱$

البته می‌توان مقادیر این توابع که حاصل یک انتگرال معین هستند را با روش‌های تقریبی انتگرال‌گیری متداول تا هر دقت لازم و مورد نیاز و تا هر چند رقم اعشار با استفاده از نرم افزارهایی مانند Maple، Matlab و S^{+} حساب کرد. بدین ترتیب با یک تابع به‌صورت جدول نمایش داده شده مواجه هستیم. در صورتی که رابطه وینزکی یا وارون آن یک رابطه ساده دستی است که بر حسب تابع معمولی ریاضی معرفی شده و می‌توان آن را در هر فرمولی که $erf(x)$ یا $erf^{-۱}(x)$ ظاهر شده جایگزین کرد. علاوه بر این محققین از زمان‌های گذشته و حتی در سال‌های اخیر نیز علاقمند به یافتن توابعی هستند که آنها را با این توابع تقریب بزنند، یا کران‌های دقیق‌تری برای آنها به‌دست آورند. برخی از این توابع عبارتند از

۱) کران‌های گوردون (۱۹۴۱):

$$\frac{z exp\{-\frac{z^{2}}{2}\}}{z^{2}+1} \leq \int_{z}^{\infty} e^{-\frac{1}{2}t^{2}} dt \leq \frac{exp\{-\frac{z^{2}}{2}\}}{z}, \quad z > 0$$

۲) کران‌های پولاک (۱۹۵۶):

$$\frac{2 exp\{-\frac{z^{2}}{2}\}}{z+(z^{2}+4)^{\frac{1}{2}}} \leq \int_{z}^{\infty} e^{-\frac{1}{2}t^{2}} dt \leq \frac{2 exp\{-\frac{z^{2}}{2}\}}{z+(z^{2}+\frac{8}{\pi})^{\frac{1}{2}}}, \quad z > 0$$

۳) کران‌های فلر (۱۹۶۸):

$$(\frac{1}{z}-\frac{1}{z^{2}})\sqrt{\frac{\pi}{2}} exp\{-\frac{z^{2}}{2}\} < 1-\Phi(z) < \frac{1}{z}\sqrt{\frac{\pi}{2}} exp\{-\frac{z^{2}}{2}\}$$

۴) کران‌های سزارگ (۱۹۹۹):

$$\frac{2 exp\{-\frac{z^{2}}{2}\}}{z+\left(z^{2}+4\right)^{\frac{1}{2}}} \leq \int_{z}^{\infty} e^{-\frac{1}{2}t^{2}} dt \leq \frac{2 exp\{-\frac{z^{2}}{2}\}}{3z+(z^{2}+8)^{\frac{1}{2}}}, \quad z > -1$$

۵) تقریب تاچر (۱۹۶۳):

$$\Phi(z) \approx \frac{exp\{2kz\}}{1+exp\{2kz\}}; \quad k=\sqrt{\frac{2}{\pi}}$$

۶) تقریب زلان و سرورو (۱۹۶۴):

$$\Phi(z) \approx 1-(0/4361t-0/1201t^{2}+0/9373t^{3})\frac{1}{\sqrt{2\pi}} exp\{-\frac{z^{2}}{2}\}$$

که در آن $t=\frac{1}{1+0/33267z}$ است.

۷) تقریب پیج (۱۹۷۷):

$$\Phi(z) \approx \frac{1}{2}(1 + \tanh y), \quad y = \sqrt{\frac{2}{\pi}}z(1 + 0.044715z^2)$$

۸) تقریب هاماکار (۱۹۷۸):

$$\Phi(z) \approx 1 - \frac{1}{2}\{1 - \sqrt{1 - e^{-y^2}}\}, \quad y = 0.0806z(1 - 0.018z^2)$$

۹) تقریب ابرناتی (۱۹۸۸):

$$\Phi(z) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{2^n n!(2n+1)}, \quad z > 0$$

۱۰) تقریب لین (۱۹۸۹):

$$\Phi(z) \approx 1 - \frac{1}{2}\{exp\{-0.717z - 0.416z^2\}\}, \quad z > 0$$

۱۱) تقریب لین (۱۹۹۰):

$$\Phi(z) \approx 1 - \frac{1}{1 + e^y}, \quad y = 4.2\pi(\frac{z}{9 - z}), \quad z > 0$$

۱۲) تقریب بگبی (۱۹۹۵):

$$\Phi(z) \approx \frac{1}{2} + \frac{1}{2}(1 - \frac{1}{30}(7e^{-\frac{1}{2}z^2} + 16e^{-(2-\sqrt{2})z^2} + (7 + \frac{\pi z^2}{4})e^{-z^2}))^{\frac{1}{2}}, \quad z > 0$$

۱۳) تقریب بیرک (۲۰۰۱):

$$\Phi(z) \approx 1 - \frac{z^2 + 5.57519z + 12.77436}{z^3\sqrt{2\pi} + 14.38718z^2 + 31.53531z + 25.548} exp\{-\frac{z^2}{2}\}$$

۱۴) تابع توزیع تجمعی لوژستیک استاندارد:

$$\Phi(z) \approx \frac{1}{1 + exp\{\frac{-\pi z}{\sqrt{3}}\}}, \quad z > 0$$

۱۵) تقریب آمیت چادهاری (۲۰۱۴):

$$\Phi(z) \approx 1 - \frac{1}{\sqrt{2\pi}} \frac{1}{0.226 + 0.64z + 0.33\sqrt{z^2 + 3}} exp\{-\frac{z^2}{2}\}, \quad z > 0$$

در جداول ۱ و ۲ به‌ترتیب ماکسیمم قدر مطلق خطا و میانگین قدر مطلق خطاها برای برخی روش‌های تقریب تابع توزیع نرمال استاندارد ارائه شده است (چادهاری، ۲۰۰۷). به‌وضوح دیده می‌شود با توجه به اینکه در بازه‌های مختلف دقت روش‌های تقریب متفاوت می‌باشد لذا بهترین تقریب به‌طور یکنواخت وجود ندارد. در این مقاله رابطه تقریبی سرگئی وینزکی با یک روش جدید در بخش ۲ اثبات می‌شود سپس در بخش ۳ مقدار خطا را در هر نقطه تعیین و نمودار آن رسم شده است در بخش ۴ به بحث و نتیجه‌گیری پرداخته خواهد شد.

۲ اثبات رابطه تقریبی سرگئی وینزکی

در این بخش ابتدا تعریف، لم و نتایج لازم برای اثبات رابطه تقریبی سرگئی وینزکی ارائه می‌شوند.

تعریف ۱ : نگارش $f(x) = O(g(x)),\ x \in D$ بدان معنا است که ثابتی مانند A وجود دارد، به قسمی که $|f(x)| \leq A|g(x)|,\ x \in D$.

جدول ۱: ماکسیمم قدر مطلق خطا برای روش‌های مختلف تقریب

روش تقریبی	$۰ < z < ۱$	$۱ < z < ۳$	$۳ < z < ۴$
تاچر	$۹/۹۱۹ \times ۱۰^{-۳}$	$۱/۷۶۷ \times ۱۰^{-۲}$	$۶/۹۱۲ \times ۱۰^{-۲}$
زلان و سرورو	$۱/۱۲۰ \times ۱۰^{-۵}$	$۱/۰۹۵ \times ۱۰^{-۵}$	$۴/۹۹۰ \times ۱۰^{-۶}$
پاج	$۱/۵۳۰ \times ۱۰^{-۵}$	$۱/۷۹۱ \times ۱۰^{-۴}$	$۱/۳۷۳ \times ۱۰^{-۴}$
هاماکار	$۶/۲۲۹ \times ۱۰^{-۴}$	$۳/۸۵۲ \times ۱۰^{-۴}$	$۲/۸۰۰ \times ۱۰^{-۶}$
لین (۸۹)	$۶/۵۸۵ \times ۱۰^{-۳}$	$۲/۳۷۴ \times ۱۰^{-۳}$	$۲/۶۹۰ \times ۱۰^{-۵}$
لین (۹۰)	$۶/۶۸۸ \times ۱۰^{-۳}$	$۲/۵۳۸ \times ۱۰^{-۳}$	$۱/۲۲۰ \times ۱۰^{-۵}$
بگبی	$۳/۰۳۰ \times ۱۰^{-۵}$	$۲/۹۶۰ \times ۱۰^{-۵}$	$۲/۷۱۰ \times ۱۰^{-۶}$
بیرک	$۱/۱۸۵ \times ۱۰^{-۵}$	$۱/۸۷۳ \times ۱۰^{-۵}$	$۲/۰۵۱ \times ۱۰^{-۶}$
لوژستیک	$۲/۲۶۶ \times ۱۰^{-۲}$	$۱/۸۴۶ \times ۱۰^{-۲}$	$۲/۰۶۳ \times ۱۰^{-۳}$

جدول ۲: میانگین قدر مطلق خطا برای روش‌های مختلف تقریب در بازه ۴ − ۰

روش تقریبی	میانگین قدر مطلق خطا
تاچر	$۸/۵۹۲ \times ۱۰^{-۳}$
زلان و سرورو	$۵/۹۸۰ \times ۱۰^{-۶}$
پاج	$۹/۴۷۰ \times ۱۰^{-۵}$
هاماکار	$۱/۶۸۲ \times ۱۰^{-۴}$
لین (۸۹)	$۱/۳۴۲ \times ۱۰^{-۳}$
لین (۹۰)	$۱/۳۶۵ \times ۱۰^{-۳}$
بگبی	$۱/۱۶۰ \times ۱۰^{-۵}$
بیرک	$۶/۹۲۱ \times ۱۰^{-۶}$
لوژستیک	$۷/۳۱۱ \times ۱۰^{-۳}$

به سادگی ثابت می‌شود، اگر $\lim_{x\to\infty}\frac{f(x)}{g(x)}$ متناهی باشد، آنگاه
$f(x)=O(g(x))$

همچنین هرگاه $x\to\infty$، آنگاه

$$CO(x^n)=O(x^n) \tag{۳}$$

$$O(x^n)+O(x^m)=O(x^n) \quad n>m \tag{۴}$$

$$O(x^n)O(x^m)=O(x^{n+m}) \tag{۵}$$

لم ۱ : برای مقادیر مثبت و بزرگ x داریم

$$erf(x)=1-\frac{e^{-x^2}}{\sqrt{\pi}x}+O(\frac{e^{-x^2}}{x^3})$$

برهان : از رابطه (۱) داریم

$$erf(x)=\frac{2}{\sqrt{\pi}}\int_0^\infty e^{-t^2}dt-\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt \tag{۶}$$

مقدار انتگرال اول رابطه (۶) برابر $erf(\infty)=1$ است، با اعمال تغییر متغیر $t=v^2$ در انتگرال دوم، خواهیم داشت

$$erf(x)=1-\frac{1}{\sqrt{\pi}}\int_{x^2}^\infty\frac{e^{-t}}{\sqrt{t}}dt$$

حال با روش جزء به جزء به دست می‌آوریم

$$erf(x)=1-\frac{1}{\sqrt{\pi}}(\frac{e^{-x^2}}{x}-\frac{1}{2}\int_{x^2}^\infty\frac{e^{-t}}{t^{\frac{3}{2}}}dt) \tag{۷}$$

ولی به ازای $t>x^2$، $\frac{1}{t^{\frac{3}{2}}}<\frac{1}{x^3}$ بنابراین

$$0<\int_{x^2}^\infty\frac{e^{-t}}{t^{\frac{3}{2}}}dt<\frac{1}{x^3}\int_{x^2}^\infty e^{-t}dt=\frac{e^{-x^2}}{x^3}$$

در نتیجه

$$\frac{1}{2\sqrt{\pi}}\int_{x^2}^\infty\frac{e^{-t}}{t^{\frac{3}{2}}}dt=O(\frac{e^{-x^2}}{x^3}) \tag{۸}$$

باتوجه به (۷) و (۸) لم ۱ ثابت می‌شود.

تذکر ۱ : از لم ۱ نتیجه می‌گردد که برای مقادیر بزرگ x داریم:

$$erf(x) \cong 1 + O(\frac{e^{-x^2}}{x}) \tag{9}$$

طرفین رابطه (۹) را به توان دو می‌رسانیم با استفاده از روابط (۳)، (۴) و (۵) خواهیم داشت:

$$(erf(x))^2 \cong 1 + O(\frac{e^{-x^2}}{x}) \tag{10}$$

لم ۲:

$$erf(x))^2 = \frac{4}{\pi}(x^2 - \frac{2x^4}{3} + O(x^6)) \tag{11}$$

برهان : با استفاده از بسط $e^{-t} = 1 - \frac{t}{1!} + \frac{t^2}{2!} - \cdots$ می‌توان نوشت

$$\begin{aligned} erf(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \\ &= \frac{2}{\sqrt{\pi}} \int_0^x (1 - \frac{t^2}{1} + \frac{t^4}{2!} - \cdots) dt \\ &= \frac{2}{\sqrt{\pi}} (x - \frac{x^3}{3} + \frac{x^5}{10} - \cdots) \end{aligned}$$

از توان دوم طرفین این عبارت نتیجه حاصل می‌شود.

لم ۳:

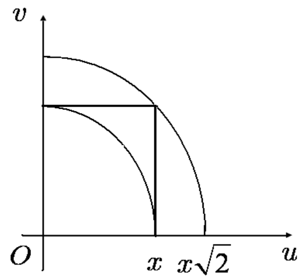
$$1 - e^{-x^2} < (erf(x))^2 < 1 - e^{-2x^2} \tag{12}$$

برهان : ابتدا فرض می‌کنیم $x \geq 0$، با در نظر گرفتن

$$(erf(x))^2 = \frac{4}{\pi}(\int_0^x e^{-u^2} du)(\int_0^x e^{-v^2} dv) = \frac{4}{\pi} \int_0^x \int_0^x e^{-(u^2+v^2)} du dv$$

و تعریف $D = \{(u,v) | 0 < u < x,\ 0 < v < x\}$، می‌توان نوشت:

$$(erf(x))^2 = \frac{4}{\pi} \int\int_D e^{-(u^2+v^2)} du dv$$



شکل ۱: نواحی انتگرال‌گیری

نواحی $D_۱$ و $D_۲$ را به‌صورت

$$D_۱ = \{(u,v)|۰ < u^۲ + v^۲ < x^۲, ۰ < u < x, ۰ < v < x\}$$
$$D_۲ = \{(u,v)|۰ < u^۲ + v^۲ < ۲x^۲, ۰ < u < x\sqrt{۲}, ۰ < v < x\sqrt{۲}\}$$

تعریف می‌کنیم. با توجه به شکل ۱ به روشنی $D_۱ \subset D \subset D_۲$ است. چون $e^{-(u^۲+v^۲)} > ۰$، می‌توان نوشت

$$\frac{۴}{\pi}\int\int_{D_۱} e^{-(u^۲+v^۲)}dudv < \frac{۴}{\pi}\int\int_{D} e^{-(u^۲+v^۲)}dudv < \frac{۴}{\pi}\int\int_{D_۲} e^{-(u^۲+v^۲)}dudv$$

با استفاده از تبدیل مختصات قطبی، انتگرال‌های فوق به‌صورت زیر ساده می‌شوند.

$$\frac{۴}{\pi}\int\int_{D_۱} e^{-(u^۲+v^۲)}dudv = \frac{۴}{\pi}\int_۰^{\frac{\pi}{۲}}\int_۰^x re^{-r^۲}drd\theta = ۱ - e^{-x^۲}$$

و به طور مشابه $\frac{۴}{\pi}\int\int_{D_۲} e^{-(u^۲+v^۲)}dudv = ۱ - e^{-۲x^۲}$ است. چون سه جمله نابرابری (۱۲) توابعی زوج هستند به‌ازای $x < ۰$ هم برقرار هستند به این ترتیب نامساوی (۱۲) ثابت شده است.

قضیه ۱ : رابطه تقریبی (رابطه سرگئی وینزکی)

$$erf(x) \cong [۱ - exp(-x^۲\frac{\frac{۴}{\pi} + ax^۲}{۱ + ax^۲})]^{\frac{۱}{۲}};\quad x \geq ۰ \qquad (۱۳)$$

را برای محاسبه تابع خطا داریم، که در آن $a = \frac{۸(\pi-۳)}{۳\pi(۴-\pi)} \cong \frac{۸۸۸۷}{۶۳۴۷۳} \cong ۰/۱۴۰۰۱$.

برهان : از نابرابری‌های (۱۲) نتیجه می‌شود که تابع $k = k(x)$ وجود دارد به طوری که $۱ < k(x) < ۲$ و $(erf(x))^۲ = ۱ - e^{-k(x)x^۲}$. چون $(erf(x))^۲$ تابعی زوج است، بنابراین $k(x)$ هم تابعی زوج است. از بین توابع تقریبی برای $k(x)$ تابع گویای

$$k(x) \cong \frac{Ax^۲ + B}{Cx^۲ + D} = \frac{\frac{A}{D}x^۲ + \frac{B}{D}}{\frac{C}{D}x^۲ + ۱}$$

یعنی $k(x) \cong \frac{ax^۲+b}{cx^۲+۱}$ را انتخاب می‌کنیم (با این انتخاب این فرم استخراج x بر حسب $erf(x)$ نیز امکان پذیر است). پس

$$(erf(x))^۲ = ۱ - e^{-\frac{ax^۲+b}{۱+cx^۲}x^۲} \qquad (۱۴)$$

حال اعداد a، b و c را چنان انتخاب می‌کنیم که

الف) وقتی $x \to \infty$ طرفین (۱۴) به‌طور مجانبی به هم نزدیک باشند.

ب) این نزدیکی برای مقادیر کوچک x هم برقرار باشد به عبارت دیگر در بسط طرفین (۱۴) برحسب توان های اولیه x با هم برابر باشند.

اعمال شرط الف: می‌توان نوشت

$$\frac{(ax^۲ + b)x^۲}{cx^۲ + ۱} = \frac{a}{c}x^۲ + \frac{(b - \frac{a}{c})x^۲}{cx^۲ + ۱}$$

در نتیجه

$$\lim_{x\to\infty} \frac{(b - \frac{a}{c})x^۲}{cx^۲ + ۱} = \frac{۱}{c}(b - \frac{a}{c})$$

پس وقتی $x \to \infty$ از (۱۴) خواهیم داشت

$$۱ - e^{-\frac{(ax^۲+b)x^۲}{cx^۲+۱}} = ۱ - e^{-\frac{a}{c}x^۲}e^{-\frac{(b-\frac{a}{c})x^۲}{cx^۲+۱}} = ۱ + O(e^{-\frac{a}{c}x^۲}) \qquad (۱۵)$$

حال نزدیکترین وضعیت عبارت های (۱۰) و (۱۵) به هم وقتی رخ می‌دهد که $\frac{a}{c} = ۱$ یعنی $a = c$، و با توجه به (۱۴) داریم

$$(erf(x))^۲ = ۱ - e^{-\frac{b+ax^۲}{۱+ax^۲}x^۲} \qquad (۱۶)$$

اعمال شرط ب: از سری $e^{-t} = ۱ - \frac{t}{۱!} + \frac{t^۲}{۲!} - \cdots$ و رابطه (۱۶) می‌توان نتیجه گرفت

$$\begin{aligned}(erf(x))^۲ &= ۱ - [۱ - \frac{۱}{۱!}(\frac{b+ax^۲}{۱+ax^۲}x^۲) + \frac{۱}{۲!}(\frac{b+ax^۲}{۱+ax^۲}x^۲)^۲ - \cdots] \\ &= \frac{۱}{۱+ax^۲}(bx^۲+ax^۴) - \frac{۱}{۲}\frac{۱}{(۱+ax^۲)^۲}(bx^۲+ax^۴)^۲ \\ &+ \cdots \end{aligned} \tag{۱۷}$$

می‌دانیم که با بسط طرفین رابطه (۱۷) بر حسب توان‌های x لازم است ضرائب مربوط به توان‌های اولیه x با هم برابر باشند. حال طرفین این رابطه را بسط می‌دهیم، با توجه به زوج بودن توان‌های x و اینکه به دو معادله روی a و b نیاز داریم و ثابت‌های دو طرف (۱۷) هم حذف می‌شوند کافیست که ضراپب $x^۲$ و $x^۴$ را مساوی هم قرار دهیم. پس عملا به بسط‌های طرفین (۱۷) تا توان چهار نیاز داریم. حال با استفاده از سری هندسی $\frac{۱}{۱+ax^۲} = ۱ - ax^۲ + a^۲x^۴ - \cdots$ و رابطه (۱۷) می‌توان نتیجه گرفت

$$\begin{aligned}(erf(x))^۲ &= (۱ - ax^۲ + \cdots)(bx^۲ + ax^۴) \\ &- \frac{۱}{۲}(۱+\cdots)(b^۲x^۴+\cdots) + \cdots \\ &= bx^۲ + (-\frac{b^۲}{۲} - ab + a)x^۴ + \cdots \end{aligned} \tag{۱۸}$$

از مساوی قرار دادن ضرائب $x^۲$ و $x^۴$ در روابط (۱۱) و (۱۸) دستگاه معادلات

$$\begin{cases} b = \frac{۴}{\pi} \\ -\frac{b^۲}{۲} - ab + a = -\frac{۸}{۳\pi} \end{cases}$$

حاصل می‌شود. به این ترتیب $a = \frac{۸(\pi-۳)}{۳\pi(۴-\pi)}$ و $b = \frac{۴}{\pi}$ به‌دست می‌آید.

۳ خطای تقریب تابع توزیع نرمال استاندارد با رابطه تقریبی سرگئی وینزکی

چون تابع خطا فرد است، بنابراین $erf(-x) = -erf(x)$، از این‌رو رابطه تقریبی سرگئی وینزکی را می‌توان به‌صورت

$$erf(x) \cong sgn(x)[۱ - exp(-x^۲\frac{\frac{۴}{\pi}+ax^۲}{۱+ax^۲})]^{\frac{۱}{۲}} \tag{۱۹}$$

نوشت. خطای تقریب تابع خطا با رابطه (۱۹) در نزدیکی صفر با مقادیر بزرگ $|x|$ ناچیز بوده و خطای کمتر از ۰/۰۰۰۳۵ است و به روش سعی و خطا می‌توان نشان داد که در صورت استفاده از مقدار $a = ۰/۱۴۷$ به‌جای $a = ۰/۱۴$ خطای آن به کمتر از ۰/۰۰۰۱۲ کاهش یابد. حال با فرض $x > ۰$، رابطه (۱۹) را بر حسب $x^۲$ حل می‌کنیم

$$exp(-x^۲\frac{\frac{۴}{\pi}+ax^۲}{۱+ax^۲}) = ۱-(erf(x))^۲$$

یا

$$x^۲\frac{\frac{۴}{\pi}+ax^۲}{۱+ax^۲} = -\ln[۱-(erf(x))^۲]$$

با قرار دادن $t = \ln[۱-(erf(x))^۲]$ داریم $ax^۴+\frac{۴}{\pi}x^۲ = -t(۱+ax^۲)$ یا $ax^۴+(\frac{۴}{\pi}+at)x^۲+t=۰$ داریم

$$x^۲ = \frac{-(\frac{۴}{\pi}+at)\pm\sqrt{(\frac{۴}{\pi}+at)^۲-۴at}}{۲a}$$

که جواب قابل قبول عبارت است از

$$x^۲ = -(\frac{۲}{\pi a}+\frac{t}{۲})+\sqrt{(\frac{۲}{\pi a}+\frac{t}{۲})-\frac{t}{a}}$$

بنابراین معکوس تابع خطا عبارت است از

$$erf^{-۱}(x) \cong \sqrt{\sqrt{(\frac{۲}{\pi a}+\frac{\ln(۱-x^۲)}{۲})^۲-\frac{\ln(۱-x^۲)}{a}}-(\frac{۲}{\pi a}+\frac{\ln(۱-x^۲)}{۲})}$$

با توجه به اینکه تابع خطا فرد است، معکوس آن نیز تابعی فرد است، در نتیجه:

$$\begin{aligned}&erf^{-۱}(x) \cong sgn(x)\\&\times\sqrt{\sqrt{(\frac{۲}{\pi a}+\frac{\ln(۱-x^۲)}{۲})^۲-\frac{\ln(۱-x^۲)}{a}}-(\frac{۲}{\pi a}+\frac{\ln(۱-x^۲)}{۲})}\end{aligned}$$

حال فرض کنیم q_α نمایش صدک αام توزیع نرمال استاندارد باشد، با استفاده از رابطه (۲) داریم

$$\alpha = \frac{۱}{۲}(۱+erf(\frac{q_\alpha}{\sqrt{۲}}))$$

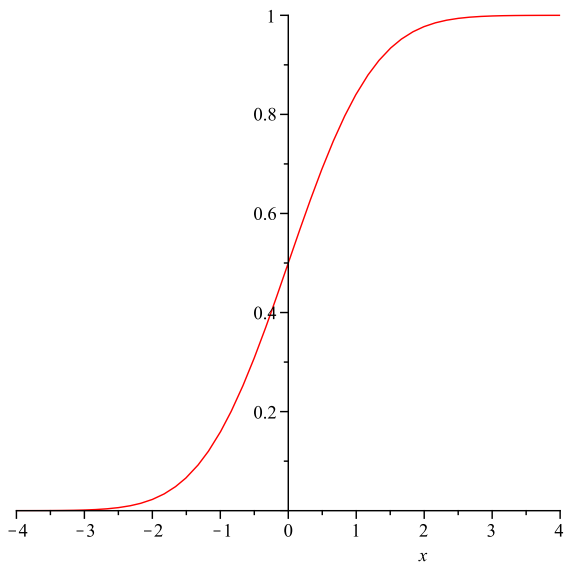
لذا صدک مذکور از رابطه

$$q_\alpha = \sqrt{۲}erf^{-۱}(۲\alpha - ۱)$$

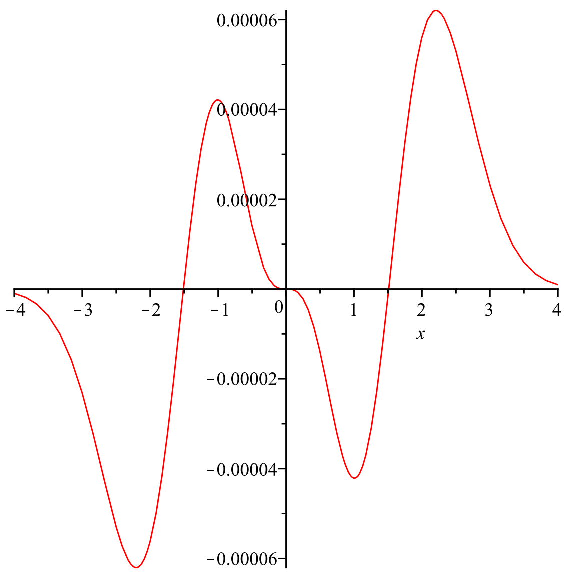
به‌دست می‌آید. با استفاده از روابط (۲) و (۱۹) می‌توان تقریب تابع $\Phi(x)$ را به کمک رابطه

$$\widehat{\Phi}(x) = \frac{۱}{۲}(۱ + sgn(x)[۱ - exp(-\frac{۱}{۲}x^۲\frac{\frac{۸}{\pi} + ax^۲}{۲ + ax^۲})]^{\frac{۱}{۲}}) \qquad (۲۰)$$

به‌دست آورد که حداکثر خطای آن ۰/۰۰۰۱۶ است. اگر در رابطه (۲۰) قرار داده شود $a = ۰/۱۴۷۳$، خطای تقریب تابع توزیع نرمال استاندارد در نزدیکی صفر یا مقادیر بزرگ $|x|$ ناچیز بوده و حداکثر خطای آن کمتر از ۰/۰۰۰۰۵۸۴ است. در شکل ۲ نمودار این تابع و در شکل ۳ نمودار خطای آن یعنی $\widehat{\Phi}(x) - \Phi(x)$ توسط نرم افزار $Maple$ رسم شده است.



شکل ۲: نمودار تقریب تابع توزیع نرمال استاندارد با رابطه (۲۰)



شکل ۳: نمودار خطا تقریب تابع توزیع نرمال استاندارد با رابطه (۲۰)

۴ بحث و نتیجه‌گیری

در این مقاله رابطه تقریبی سرگئی وینزکی برای تقریب تابع تویع نرمال استاندارد برای تخمین مساحت زیر منحنی نرمال استاندارد به روشی جدید ثابت و با روش سعی و خطا و تغییر یکی از پارامترهای آن رابطه‌ای با دقت بالاتر نسبت به رابطه تقریبی سرگئی وینزکی به دست آورده شده است. همانطور که از جداول ۱ و ۲ شکل ۳ نتیجه می‌شود دقت این رابطه در بیشتر موارد از سایر روابط بیشتر است. البته ادعا نمی‌شود که جدول توزیع احتمال نرمال باید کنار گذاشته شود اما تقریب ارائه شده می‌تواند رقیبی برای جدول توزیع احتمال نرمال باشد.

تقدیر و تشکر

نویسنده لازم می‌داند که از جناب آقای دکتر عبدالحسین هورفر استاد برجسته دانشگاه تهران، داوران و مسئولین محترم مجله علوم آماری به دلیل دقت در بازبینی مقاله و پیشنهادات ارزنده‌شان که منجر به بهبود این مقاله شده است، تشکر و

قدردانی نماید.

مراجع

Abernathy, R. W. (1988), Finding Normal Probabilities with Hand held Calculators, *Mathematics Teacher*, **81**, 651-652.

Bagby, R. J. (1995), Calculating Normal Probabilities, *American Mathematical Monthly*, **102**, 46-49.

Byrc, W. (2001), A Uniform Approximation to the Right Normal Tail Integral, *Applied Mathematics and Computation*, **127**, 365-374.

Choudhury, A., Ray, S. and Sarkar, P. (2007), Approximating the Cumulative Distribution of the Normal Distribution, *Mathematics and Journal of Statistical Research*, **41**, 59-67.

Choudhury, A. (2014), A Simple Approximation to the Area Under Standard Normal Curve, *Mathematics and Statistics*, **2**, 147-149.

Feller, W. (1968), *An Introduction to Probability Theory and its Applications*, John Wiley, New York.

Gordon, R. D. (1941), Values of Mills' Ratio of Area Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument, *Annales of Mathematical Statistics*, **12**, 364-366.

Hammakar, H. C. (1978), Approximating the Cumulative Normal Distribution and its Inverse, *Applied Statistics*, **27**, 76-77.

Lin, J. T. (1989), Approximating the Normal Tail Probability and its Inverse for Use on a Pocket Calculator, *Applied Statistics*, **38**, 69-70.

Lin, J. T. (1990), A Simpler Logistic Approximation to the Normal Tail Probability and its Inverse, *Applied Statistics*, **39**, 255-257.

Page, E. (1977), Approximations to the Cumulative Normal Function and its Inverse for Use on a Pocket Calculator, *Applied Statistics*, **26**, 75-76.

Pollak, H. O. (1956), A Remark on "Elementary Inequalities for Mills' Ratio" by Y. Komatu, *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers*, **4**, 110.

Szarek, S. J. (1999), A Nonsymmetric Correlation Inequality for Gaussian Measure, *Journal of Multivariate Analysis*, **68**, 193-211.

Tocher, K. D. (1963), *The Art of Simulation*, English University Press, London.

Winitzki, S., http://sites.google.com/site/winitzki/sergei winitzki's files.

Zelen, M. and Severo, N. C. (1964), Probability Function (pp 925-995) in Handbook of Mathematical Functions, Abramowitz, M. and Stegun, I. A., *Applied Mathematics Series*, 55.