

استنباط آماری براساس روش برآورد گر گشتاوری وزنی احتمالی تعدیل یافته

افشین فیاض موقر^{۱*}، زهره محمودی^۲

۱- استادیار دانشگاه مازندران، دانشکده علوم ریاضی، گروه آمار، بابلسر، ایران

۲- دانشجوی کارشناسی ارشد، دانشگاه مازندران، گروه آمار، بابلسر، ایران

رسید مقاله: ۲ اسفند ۱۳۹۳

پذیرش مقاله: ۴ مرداد ۱۳۹۴

چکیده

در این مقاله روش جدیدی در نظریه برآوردیابی با عنوان برآورد گر گشتاوری وزنی احتمالی تعدیل یافته (APWM)^۲ معرفی می گردد، که تعمیمی از روش گشتاوری وزنی احتمالی (PWM)^۳ می باشد. با استفاده از این روش استنباط آماری بر روی پارامتر توزیع یکنواخت $(0, \theta)$ انجام داده ایم، و نتیجه آن را با نتایج روش های دیگر برآوردیابی از جمله روش حداکثر درست نمایی (ML)^۴ که برای توزیع یکنواخت دارای کوتاه ترین فاصله اطمینان می باشد و روش گشتاوری وزنی احتمالی (PWM) مقایسه نموده ایم. نتایج بیانگر برتری APWM نسبت به PWM می باشد و در اکثر مواقع حداقل نتایج مشابه ML می دهد.

کلمات کلیدی: برآورد گر گشتاوری وزنی احتمالی، مجموع اشتیل یس تصادفی.

۱ مقدمه

برآورد پارامترهای توزیع های احتمالی و معرفی برآورد گرهای مناسب یکی از موضوع های مهم در آمار می باشد، که از جمله روش های برآوردیابی می توان به روش گشتاوری (MM)^۵، روش حداکثر درست نمایی (ML) و... اشاره کرد [۱].

براساس این برآورد گرها استنباط هایی در رابطه با پارامتر توزیع صورت می پذیرد. از جمله این استنباط ها می توان فاصله اطمینان و انجام آزمون را نام برد. مفاهیم برآورد فاصله ای و آزمون فرض ها ارتباط نزدیکی با هم دارند و هر یک از این مفاهیم می تواند منجر به معرفی مفهوم دیگر شود. همان طور که در برآورد نقطه ای معیار بررسی عملکرد روش های مختلف برآوردیابی، میانگین مربع خطا^۶ می باشد در برآورد فاصله ای برای بررسی

* عهده دار مکاتبات

آدرس الکترونیکی: a_fayyaz@umz.ac.ir

² Adjusted Probability Weighted Moments

³ Probability Weighted Moments

⁴ Maximum Likelihood

⁵ Method of Moments

⁶ Mean Square Error

عملکرد، توان^۱ مربوط به روش‌ها را معیار قرار می‌دهیم. در این مقاله به بررسی عملکرد روش‌های مختلف برآوردیابی در انجام استنباط آماری می‌پردازیم.

در بخش دوم به معرفی، روش گشتاوری وزنی احتمال (PWM) و در بخش سوم روش جدید، گشتاوری وزنی احتمالی تعدیل یافته (APWM)، خواهیم پرداخت. در بخش چهارم استنباط‌های آماری مربوط به پارامتر توزیع یکنواخت ارایه می‌گردد و مقایسه عددی مربوط به استنباط‌های مختلف در بخش پنجم آمده است.

۲ روش برآوردیابی PWM

شاید بتوان گفت که روش گشتاوری یکی از قدیمی‌ترین روش‌ها برای برآورد پارامترها می‌باشد، که توسط کارل پیرسن در سال ۱۸۹۴ معرفی گردید. در سال ۱۹۷۹ گرین وود [۲] با وزنی کردن روش گشتاوری (MM)، روش جدیدی را با عنوان روش گشتاوری وزنی احتمالی (PWM) معرفی کرد، که در واقع تعمیمی از روش گشتاوری می‌باشد.

گشتاور وزنی احتمالی یک متغیر تصادفی X ، با تابع توزیع $F_{\theta}(X) = P_{\theta}(X \leq x)$ به صورت زیر معرفی می‌گردد:

$$M_{(p,r,s)} = E_{\theta} \left[X^p \{F_{\theta}(X)\}^r \{1 - F_{\theta}(X)\}^s \right] \quad (۱)$$

که در آن p, r و s مقادیر حقیقی می‌باشند. هنگامی که معکوس تابع توزیع، $x(F)$ ، را بتوان به فرم بسته‌ای نوشت، رابطه (۱) به صورت زیر نوشته می‌شود:

$$M_{(p,r,s)} = \int_0^1 \{x(F)\}^p F^r (1-F)^s dF \quad (۲)$$

که در آن F دارای توزیع یکنواخت $(0,1)$ می‌باشد، که به این ترتیب گشتاورها وزنی احتمالی را می‌توان آسان‌تر به دست آورد. همان‌طور که مشاهده می‌شود کمیت‌های $M_{(p,.,.)}$ ($p = 1, 2, \dots$) گشتاورهای معمولی غیر مرکزی به دست می‌دهند. آماره برآوردگر پارامتر θ مشابه روش برآوردیابی گشتاوری، از برابری گشتاورهای وزنی احتمالی جامعه (۱) با مقدار به دست آمده آن براساس نمونه حاصل می‌گردد. در حالت خاص که $s = 0$ یا $r = 0$ فرض می‌گردد [۳] و برآوردگر پارامتر θ از رابطه‌های زیر حاصل می‌شود:

$$a_s = M_{(1,.,s)} = E_{\theta} \left[X \{1 - F_{\theta}(X)\}^s \right] \quad (۳)$$

یا

$$b_r = M_{(1,r,.)} = E_{\theta} \left[X \{F_{\theta}(X)\}^r \right] \quad (۴)$$

که در آن

^۱ Power

$$a_s = \frac{1}{n} \sum_{j=1}^n \frac{\binom{n-j}{r}}{\binom{n-1}{r}} x_{(j)}, \quad (5)$$

و

$$b_r = \frac{1}{n} \sum_{j=1}^n \frac{\binom{j-1}{r}}{\binom{n-1}{r}} x_{(j)} \quad (6)$$

و $x_{(j)}$ مقدار مشاهده شده $X_{(j)}$ ، j -امین آماره ترتیبی نمونه تصادفی n تایی می باشد. از این روش در برآورد پارامترهای چندین توزیع مهم از جمله گامبل، لجستیک و وایبل استفاده شده است که نتایج مطلوبی نسبت به سایر روش های برآوردیابی ارائه نموده است [۳ و ۴]. همچنین از خواص این برآوردگر می توان به مجانبی نرمال بودن آن اشاره کرد [۴].

۳ معرفی روش جدید برآوردیابی APWM

در این بخش، روشی جدید در برآوردیابی پارامتر جامعه با عنوان، برآورد گشتاوری وزنی احتمالی تعدیل یافته (APWM) معرفی می گردد که در واقع تعمیم یافته روش گشتاوری وزنی احتمالی می باشد. این ایده از روش گشتاوری تعدیل یافته (AMM)^۱، که توسط سلطانی و حومه ای [۵] در سال ۲۰۰۹ معرفی گردیده، گرفته شده است. روش جدید مشابه روش PWM می باشد، اما به جای استفاده از گشتاورهای وزنی احتمالی نمونه از مجموع جزئی استلتنجس تصادفی (RSPTS)^۲ متناظر استفاده می گردد.

فرض کنید $X_1, \dots, X_n$ یک نمونه تصادفی n تایی از جامعه ای با تابع توزیع دلخواه F_θ با پارامتر مجهول θ باشد. همان طور که در بخش ۱ به آن اشاره شد گشتاور وزنی احتمالی را می توان به صورت زیر نوشت:

$$M_{(p,r,s)} = E_\theta \left[X^p \{F_\theta(X)\}^r \{1-F_\theta(X)\}^s \right] = \int \{x(F)\}^p F^r (1-F)^s dF \quad (7)$$

بر اساس نظریه انتگرال گیری کلاسیک، انتگرال موجود در رابطه بالا را می توان از مجموع استلتنجس تصادفی بالایی (پایینی) برای $p, r, s = 0, 1, 2, \dots$ به صورت زیر تقریب زد:

$$X_{(U)(p,r,s)} = \sum_{j=1}^n X_{(j)}^p \left\{ F(X_{(j)}) \right\}^r \left\{ 1-F(X_{(j)}) \right\}^s \left[F(X_{(j)}) - F(X_{(j-1)}) \right],$$

$$(X_{(L)F} = \sum_{j=1}^n X_{(j)}^p \left\{ F(X_{(j)}) \right\}^r \left\{ 1-F(X_{(j)}) \right\}^s \left[F(X_{(j+1)}) - F(X_{(j)}) \right]) \quad (8)$$

که در آن $F(X_{(0)}) = 0$ ، $F(X_{(n+1)}) = 1$ و $X_{(1)} \leq \dots \leq X_{(n)}$ آماره متناظر با $X_1, \dots, X_n$ می باشد. در این مقاله ما برآورد پارامتر مجهول θ را از حل دستگاه غیر خطی

¹ Adjusted Method of Moments

² Random Stieltjes Partial Sums

$$M_{(p,r,s)} = X_{(U)(p,r,s)} \quad (9)$$

برآورد می کنیم.

فیاض و محمودی [۶] برآورد APWM را برای توزیع یکنواخت $(0, \theta)$ به دست آورده اند و مقایسه ای بین نتایج این روش و روش های حداکثر درست نمایی، گشتاوری تعدیل یافته و گشتاوری وزنی احتمالی برای توزیع یکنواخت از طریق مقایسه میانگین مربع خطا انجام داده اند که نشان داد روش APWM برتری هایی نسبت به بقیه روش ها دارد. برآورد APWM پارامتر توزیع یکنواخت $(0, \theta)$ به صورت زیر به دست می آید [۶]:

$$\hat{\theta}_{APWM} = \left[(r+2) \sum_{j=1}^n U_{(j)}^{(r+1)} (U_{(j)} - U_{(j-1)}) \right]^{1/(r+2)}$$

سلطانی و عبدا... نژاد دو روش استنباط آماری را بر طبق برآوردگرهای گشتاوری تعدیل یافته معرفی کرده اند [۷ و ۸].

در این مقاله هدف ما انجام استنباط آماری بر روی پارامتر مجهول توزیع یکنواخت می باشد، که در ادامه به این موضوع خواهیم پرداخت.

۴ استنباط برای پارامتر توزیع یکنواخت

سلطانی و عبدا... نژاد با استفاده از برآوردگرهای روش AMM بر روی پارامتر θ توزیع یکنواخت، استنباط آماری انجام داده اند [۷] که ما در این بخش با استفاده از برآوردگرهای APWM استنباط آماری بر روی پارامتر توزیع یکنواخت انجام خواهیم داد.

در این بخش جهت استنباط در مورد پارامتر θ ، فاصله اطمینانی را برای آن با استفاده از APWM برآوردگر ساخته و آزمون فرضی بر روی آن انجام می گردد.

۴-۱ فاصله اطمینان

فرض کنید، $U_1, \dots, U_n$ یک نمونه تصادفی از $U(0, \theta)$ با پارامتر θ باشد و $U_{(1)}, \dots, U_{(n)}$ آماره های مرتب متناظر با این نمونه تصادفی باشند.

کوتاه ترین فاصله اطمینان $100(1-\alpha)\%$ پارامتر θ ، حاصل از برآوردگر ML

$$\left(U_{(n)}, \frac{U_{(n)}}{\sqrt[n]{\alpha}} \right) \quad (10)$$

می باشد [۱].

به منظور به دست آوردن فاصله اطمینان APWM عبارت زیر را در نظر بگیرید:

$$B = \frac{\hat{\theta}_{APWM}}{\theta} = \left[(r+2) \sum_{j=1}^n \frac{U_{(j)}}{\theta} \left\{ \frac{U_{(j)}}{\theta} \right\}^r \left(\left(\frac{U_{(j)}}{\theta} \right) - \left(\frac{U_{(j-1)}}{\theta} \right) \right) \right]^{1/(r+2)}$$

¹ Shortest Confidence Interval

از آنجا که توزیع عبارت بالا مستقل از θ می باشد بنابراین به عنوان کمیت محوری^۱ در نظر گرفته می شود. بنابراین فاصله اطمینان $\%(1-\alpha)100$ پارامتر θ براساس کمیت محوری B به صورت زیر محاسبه می گردد. می دانیم که

$$P\left(B\left(\frac{\alpha}{r}\right) \leq \frac{\hat{\theta}_{APWM}}{\theta} \leq B\left(1-\frac{\alpha}{r}\right)\right) = 1-\alpha,$$

که در آن $B\left(\frac{\alpha}{r}\right)$ و $B\left(1-\frac{\alpha}{r}\right)$ چندک های متغیر تصادفی B می باشند.

در نتیجه داریم:

$$1-\alpha = P\left(B\left(\frac{\alpha}{r}\right) \leq \frac{\hat{\theta}_{APWM}}{\theta} \leq B\left(1-\frac{\alpha}{r}\right)\right) = P\left(\frac{\left[(r+r)\sum_{j=1}^n u_{(j)}^{(r+1)}(u_{(j)}-u_{(j-1)})\right]^{\frac{1}{r+r}}}{B\left(1-\frac{\alpha}{r}\right)} \leq \theta \leq \frac{\left[(r+r)\sum_{j=1}^n u_{(j)}^{(r+1)}(u_{(j)}-u_{(j-1)})\right]^{\frac{1}{r+r}}}{B\left(\frac{\alpha}{r}\right)}\right).$$

متغیر تصادفی T را به صورت زیر تعریف می کنیم:

$$T = \frac{\left[(r+r)\sum_{j=1}^n u_{(j)}^{(r+1)}(u_{(j)}-u_{(j-1)})\right]^{\frac{1}{r+r}}}{B} = \left(\frac{\sum_{j=1}^n u_{(j)}^{(r+1)}(u_{(j)}-u_{(j-1)})}{\sum_{j=1}^n V_{(j)}^{(r+1)}(V_{(j)}-V_{(j-1)})}\right)^{\frac{1}{r+r}} \quad (11)$$

که در آن $u_{(j)}$ مقدار مشاهده شده، $U_{(j)}$ ، $j=1,2,\dots,n$ و $V_{(j)} = \frac{dU_{(j)}}{\theta}$ می باشد (علامت d نشان دهنده هم توزیع بودن^۲ است).

حال فاصله اطمینان $\%(1-\alpha)100$ برای پارامتر θ را می توان به صورت

$$\left(T\left(\frac{\alpha}{r}\right), T\left(1-\frac{\alpha}{r}\right)\right)$$

به دست آورد که در آن $T(\alpha)$ چندک α ام توزیع T می باشد. در عمل چندک های T از طریق شبیه سازی به دست می آیند.

۴-۲ آزمون فرض

برای انجام آزمون فرض برای پارامتر θ ، سه نوع آزمون می توانیم انجام دهیم:

- (i) $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$,
- (ii) $H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$,

¹ Pivotal Quantity

² Identically Distributed

$$(iii) H_1: \theta \geq \theta_0 \text{ vs } H_0: \theta < \theta_0.$$

در روش ML ناحیه بحرانی مربوط به پرتوان ترین آزمون یکنواخت، فواصل راست و چپ فاصله اطمینان بیان شده در رابطه (۱۰) می باشد، یعنی

$$\left(-\infty, U_{(n)}\right) \cup \left(\frac{U_{(n)}}{\sqrt[n]{\alpha}}, \infty\right),$$

در روش APWM نیز ناحیه رد هر یک از آزمون های بالا بر حسب متغیر تصادفی T به ترتیب عبارتند از:

$$i. |T| \geq \theta,$$

$$ii. T \leq \theta,$$

$$iii. T \geq \theta.$$

بنابراین P-مقدار^۱ این آزمون ها که آن ها را با P نمایش می دهیم، برابرند با:

$$i. P = \min\{P(T \leq \theta), P(T \geq \theta)\}$$

$$ii. P = P(T \leq \theta),$$

$$iii. P = P(T \geq \theta).$$

۵ مثال عددی

در این بخش محاسبات عددی مربوط به نتایج حاصل شده در بخش ۴ براساس شبیه سازی به دست می آید. در ادامه این نتایج با روش های PWM و ML به صورت عددی مقایسه می گردد.

۵-۱ بررسی فاصله اطمینان

جهت مقایسه میانگین طول^۲ و احتمال پوشش^۳ فواصل اطمینان حاصل از برآوردگر APWM، PWM و ML (کوتاه ترین فاصله اطمینان) از شبیه سازی استفاده می شود. الگوریتم زیر مراحل شبیه سازی متغیر تصادفی T را در چهار گام نشان می دهد:

۱. تولید U_i ها از $U(\cdot, \theta)$ ، $i = 1, 2, \dots, n$

۲. تولید V_i ها از $U(\cdot, 1)$ ، $i = 1, 2, \dots, n$

۳. محاسبه متغیر T با استفاده از رابطه (۱۱)،

۴. m بار تکرار سه مرحله قبل.

جدول (۱) نتایج مربوط به این مقایسه به ازای n های متفاوت، $r = 2$ ، $m = 1000$ و $\alpha = 0.05$ می باشد، که بیانگر آن است که احتمال پوشش برای هر اندازه نمونه، نزدیک به سطح ۰/۹۵ می باشد و روش APWM به خوبی ML عمل کرده است.

¹ P-value

² Length Average

³ Coverage Probability

کمترین احتمال پوشش در بین این سه روش متعلق به روش PWM می‌باشد. همچنین برای هر n ، APWM دارای میانگین طول کمتری نسبت به PWM می‌باشد.

جدول ۱. احتمال پوشش (C) و میانگین طول (L) فاصله اطمینان دو طرفه ۰/۹۵ برای پارامتر توزیع یکنواخت با $\theta = 2$ با توجه به حجم های نمونه مختلف

نمونه (n) روش	۵	۱۰	۳۰	۵۰	۷۰	۱۰۰	۵۰۰
APWM							
C	۰/۹۵	۰/۹۵	۰/۹۶	۰/۹۵	۰/۹۶	۰/۹۵	۰/۹۵
L	۱/۲۱۸۶	۰/۷۵۳۰	۰/۲۲۸۲	۰/۱۷۳۷	۰/۱۱۰۳	۰/۰۷۵۵	۰/۰۴۳۲
(PWM)							
C	۰/۸۷	۰/۸۸	۰/۹۲	۰/۹۵	۰/۹۴	۰/۹۵	۰/۹۵
L	۲/۲۵۸۵	۱/۲۳۵۷	۰/۶۱۶۶	۰/۴۳۲۲	۰/۳۷۲۷	۰/۲۶۸۴	۰/۱۱۳۳
(ML)							
C	۰/۹۵	۰/۹۶	۰/۹۵	۰/۹۶	۰/۹۶	۰/۹۵	۰/۹۵
L	۱/۴۷۲۳	۰/۶۷۶۴	۰/۲۰۴۸	۰/۱۲۲۲	۰/۰۸۵۶	۰/۰۵۹۹	۰/۰۳۲۱

۵-۲ نتایج مربوط به آزمون فرض

یک مطالعه شبیه سازی جهت مقایسه اندازه^۱ و توان آزمون براساس برآوردگر APWM، PWM و ML صورت گرفته است. بدین منظور نمونه هایی با اندازه های نمونه متفاوت (n) از توزیع $U(0, \theta)$ تولید کرده و با ۵۰۰۰ بار تکرار این عمل، توان و اندازه آزمون را به دست می آوریم. نتایج عددی برای $\theta = r = 2$ در جداول ۲ و ۳ گردآوری شده اند.

در بیشتر مواقع، اندازه APWM در بین سه روش دیگر بهتر و همیشه نزدیک به سطح $\alpha = 0.05$ می باشد (جدول ۲). جدول (۳) توان آزمون را برای فرض (i) نشان می دهد. همان طور که مشاهده می گردد در سطح $\alpha = 0.05$ ، APWM دارای توان بزرگتری می باشد و PWM نتایج ضعیف تری را نشان می دهد، در حالی که APWM و ML به یکدیگر نزدیک می باشند.

جدول ۲. اندازه آزمون در آزمون دو طرفه $\alpha = 0.05$ توزیع یکنواخت به ازای $\theta = 2$ برای حجم های نمونه متفاوت

n روش	۵	۱۰	۳۰	۵۰	۷۰	۱۰۰	۵۰۰
APWM	۰/۰۵	۰/۰۵	۰/۰۴	۰/۰۵	۰/۰۴	۰/۰۵	۰/۰۵
PWM	۰/۱۳	۰/۱۲	۰/۰۸	۰/۰۵	۰/۰۶	۰/۰۵	۰/۰۵
ML	۰/۰۵	۰/۰۴	۰/۰۵	۰/۰۴	۰/۰۴	۰/۰۵	۰/۰۵

¹ Size

جدول ۳. توان آزمون در آزمون دو طرفه $\alpha = 0.05$ توزیع یکنواخت به ازای $\theta = 2$ برای θ_0 ها و حجم های نمونه متفاوت

روش	نمونه (n)	۵	۱۰	۳۰	۵۰	۷۰	۱۰۰	۵۰۰
$(\theta_0 = 1/5)$								
APWM		۰/۸۰	۰/۹۶	۱	۱	۱	۱	۱
PWM		۰/۷۷	۰/۸۸	۱	۱	۱	۱	۱
ML		۰/۷۳	۰/۹۵	۱	۱	۱	۱	۱
$(\theta_0 = 1/15)$								
APWM		۰/۵۸	۰/۷۷	۰/۹۸	۱	۱	۱	۱
PWM		۰/۴۹	۰/۵۷	۰/۷۴	۰/۸۳	۰/۹۵	۰/۹۹	۱
ML		۰/۶۲	۰/۸۴	۰/۹۸	۱	۱	۱	۱
$(\theta_0 = 2)$								
APWM		۰/۱۷	۰/۳۷	۰/۷۳	۰/۹۴	۰/۹۶	۱	۱
PWM		۰/۳۱	۰/۳۴	۰/۳۵	۰/۴۱	۰/۴۸	۰/۵۴	۰/۸۹
ML		۰/۲۷	۰/۵۰	۰/۷۸	۰/۹۶	۰/۹۵	۱	۱

۶ نتیجه گیری

در این مقاله ما بر پایه روش APWM یک فاصله اطمینان معرفی کردیم و از طریق شبیه سازی احتمال پوشش و میانگین طول روش APWM را با روش های ML و PWM مورد بررسی قرار دادیم. برای حجم های متفاوت نمونه کوتاه ترین فاصله اطمینان مربوط به روش APWM بوده است. علاوه بر این احتمال پوشش به سطح اطمینان ۰/۹۵ نزدیک می باشد. همچنین با توجه به نتایج عددی بخش آزمون فرض ها عملکرد بهتر روش APWM را نسبت به دو روش تحت بررسی نشان می دهد. بنابراین با توجه به نتایج عددی مشاهده شده، می توان گفت که روش APWM به همان خوبی روش ML عمل می کند. این در حالی است که روش PWM ضعیف ترین نتایج را به دست می دهد.

منابع

[۶] فیاض موقر، ا.، محمودی، ز.، (۱۳۹۳). روش نوین برآوردیابی: برآوردگرهای گشتاوری با وزن احتمالی تعدیل یافته، دوازدهمین کنفرانس آمار ایران، کرمانشاه.

- [1] Casella, G., Berger, R. L., (1989). Statistics Inference, Wadsworth, Pacific Grove, CA.
- [2] Greenwood, J. A., Landwehr, J. M., Matalas, N. C., Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. Water Resources Research, 15, 1049-1054.
- [3] Landwehr, J. M., Matalas, N. C., Wallis, J. R., (1979). Probability weighted moments compared with some traditional in estimation gumbel parameters and quantiles, Water Resources Research, 15, 1055-1064
- [4] Hosking, J. R. M., Wallis, J. R., Wood, E. F., (1985). Estimation of the generalized extreme value distribution by the method of probability weighed moments. Technometrics, 27, 251-261.
- [5] Soltani, A. R., Homei, H., (2009). A generalization for two-sided power distribution and adjusted method of moments. Statistics, 43, 611-620.

- [7] Soltani, A. R., Abdollahnezhad, K., (2011). On adjusted method of moments estimators on uniform distribution samples, *Metron*, 70, 27-40.
- [8] Soltani, A. R., Abdollahnezhad, K., (2012). A new computational test procedure adjusted method of moments, *Journal of Applied Probability and Statistics*, 7, 9-18.