



## تعیین حجم نمونه‌ی بیزی با کمترین هزینه بر اساس تابع زیان پیشگیرانه

مسعود قاسمی بهجانی<sup>۱</sup>، حسن زارعی

دانشگاه سیستان و بلوچستان

چکیده: در این مقاله با استفاده از تابع زیان نامتقارن پیشگیرانه، شیوه‌ی تعیین حجم نمونه به روش بیز را به قسمی بیان می‌کنیم که هزینه‌ی نمونه‌گیری مینیمم شود و در ادامه با در نظر گرفتن توزیع گاما به‌عنوان توزیع پیشین و تابع هزینه‌ی لیندلی و با مینیمم کردن هزینه‌ی نمونه‌گیری حجم نمونه‌ی مطلوب را به روش بیز محاسبه می‌کنیم.

واژه‌های کلیدی: برآوردگر بیز، تابع زیان پیشگیرانه، توزیع پسین و توزیع گاما-گاما.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62C10، 62Cxx، 62 – XX

### ۱ مقدمه

محاسبه‌ی حجم نمونه در مسائل عملی سوال مهمی است که مستقیماً با هزینه‌ی مطالعه و گردآوری داده‌ها ارتباط دارد. در هر پژوهش، دانستن تعداد مشاهدات (اندازه‌ی نمونه) به‌طوری که هزینه‌ی کلی طرح نیز مینیمم شود مسئله‌ای است که از اهمیت زیادی برخوردار است. هزینه‌ی کلی شامل زیان حاصل شده از تصمیم‌گیری و هزینه‌ی آنالیز و هدایت آزمایش‌های انجام شده است. این هزینه به عامل‌های زیادی بستگی دارد. از جمله‌ی عوامل مهم، یکی اندازه‌ی نمونه‌ی نهایی و دیگری روش جمع‌آوری داده‌ها است. برای داشتن زیان کمتر باید آزمایش‌های بیشتری انجام شود که این امر باعث افزایش هزینه‌ی آزمایش‌ها می‌شود. اولین مقاله بر اساس نظریه‌ی تصمیم برای تعیین اندازه‌ی نمونه توسط گروندی<sup>۱</sup> و همکاران (۱۹۵۶) بر پایه‌ی تابع مطلوبیت ارائه شد، این روش توسط همیلتون (۱۹۶۰)<sup>۲</sup> دوباره مطرح شده و توجه بسیاری از پژوهشگران را به خود جلب کرد. استالارد<sup>۳</sup> (۱۹۹۶) با بکارگیری این روش در فاز دوم

<sup>۱</sup> مسعود قاسمی بهجانی . mexx\_2592@yahoo.com

A-10-424-1

Grundy<sup>۱</sup>

Hamilton<sup>۲</sup>

Stalard<sup>۳</sup>

آزمایه‌های بالینی، اندازه‌ی نمونه را با ماکسیم‌سازی تابع مطلوبیتی خاص بدست آورد. پزشک و گیتینز<sup>۴</sup> (۱۹۹۹) مسئله‌ی تعیین اندازه‌ی نمونه را با استفاده از روش بیز برای یک آزمایه‌ی بالینی مطرح و آن را برای حالتی خاص حل کردند. اوهاگان و استیونس<sup>۵</sup> (۲۰۰۱) یک تابع هدف شامل هزینه و سود انجام یک آزمایه‌ی تصادفی را مطرح نموده و با بهینه‌سازی آن، اندازه‌ی نمونه‌ی آزمایه‌ی تصادفی را به روش بیزی به دست آوردند. همچنین پزشک (۲۰۰۳) مروری بر روش‌های مختلف بیزی برای تعیین اندازه‌ی نمونه‌ی آزمایه‌های تصادفی ارائه داد. در این مقاله با در نظر گرفتن تابع هزینه‌ی لیندلی و تابع زیان پیشگیرانه به روش بیز به دنبال تعیین حجم نمونه می‌باشیم به طوری که هزینه‌ی نمونه‌گیری مینیمم شود.

## ۲ تابع هزینه

تابع هزینه‌ی خطی ارائه شده توسط لیندلی (۱۹۷۲)<sup>۶</sup> به صورت  $C(n) = c_0 + cn$  ;  $n > 0$  می‌باشد که در آن  $c$  هزینه‌ی هر واحد نمونه‌گیری است و  $c_0$  هزینه‌ی زیر ساختی نمونه‌گیری برای کلیت نمونه‌گیری یا هر هزینه‌ی دیگری که در نمونه‌گیری است می‌باشد و  $C(0) = 0$ ، یعنی اگر تعداد نمونه صفر باشد، هزینه‌ی صورت نگرفته و مقدار هزینه صفر است. اگر طرح و برنامه‌ای برای نمونه‌گیری وجود نداشته باشد بوضوح هزینه‌های مربوط به روش‌های نمونه‌گیری وجود ندارد اما بدیهی است که در این حالت دقت نمونه‌گیری کاهش می‌یابد. ما با مینیم کردن هزینه‌ی نمونه‌گیری حجم نمونه‌ی مطلوب را برآورد خواهیم کرد. فرض کنید  $a$  یک مقدار بر حسب واحد های  $C(n)$  باشد که آن را برای کاهش مقدار ریسک پسین در نظر می‌گیریم. بنابراین هزینه‌ی کل به صورت زیر است:

$$TC(n) = c_0 + cn + a(PR) . \quad (1.2)$$

از طرفی می‌توانیم این تابع هزینه را به صورت زیر بنویسیم:

$$TC'(n) = c'_0 + c'n + PR ,$$

که در آن

$$c' = \frac{c}{a}, c'_0 = \frac{c_0}{a}, TC'(n) = \frac{TC(n)}{a} .$$

پس می‌توانیم بگوییم مینیم  $C(n)$  با مینیم  $C'(n)$  معادل است. و برای راحتی کار  $a$  را برابر با یک قرار می‌دهیم.

Gitins and Pezeshk<sup>۴</sup>

Stivence and O'hagan<sup>۵</sup>

Lindley<sup>۶</sup>

### ۳ توزیع گاما-گاما

برناردو<sup>۷</sup> (۱۹۹۴) توزیع گاما-گاما را بصورت زیر تعریف کرد:

متغیر تصادفی  $X$  دارای توزیع گاما-گاما با پارامترهای  $\alpha$ ،  $\beta$  و  $n$  است اگر تابع چگالی آن به صورت زیر باشد:

$$g_g(x; \alpha, \beta, n) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+n)}{\Gamma(n)} \frac{x^{n-1}}{(\beta+x)^{\alpha+n}} ; \quad x > 0 .$$

که در آن  $\alpha > 0$ ،  $\beta > 0$  و  $n$  عدد صحیح مثبت است. اگر  $X$  دارای توزیع گاما با پارامترهای  $n$  و  $\lambda$  باشد آن را با نماد  $g_a(x; n, \lambda)$  نشان می‌دهیم، در این صورت به راحتی می‌توان نشان داد که توزیع گاما-گاما از ترکیب دو توزیع گاما به صورت زیر به دست می‌آید:

$$g_g(x; \alpha, \beta, n) = \int_0^\infty g_a(x; n, \lambda) g_a(\lambda; \alpha, \beta) d\lambda .$$

مثال: اگر  $X_1, X_2, \dots, X_n$  یک نمونه‌ی تصادفی به حجم  $n$  از توزیع نمایی با پارامتر  $\lambda$  باشد و توزیع پیشین را گاما با پارامترهای  $\alpha$  و  $\beta$  در نظر بگیریم بدیهی است که توزیع پسین، گاما با پارامترهای  $\alpha+n$  و  $\beta+S$  می‌باشد و  $S = \sum X_i$  دارای توزیع گاما-گاما با پارامترهای  $\alpha$ ،  $\beta$  و  $n$  است که به صورت زیر بدست می‌آید:

$$\begin{aligned} p(s) &= \int_\lambda p(s|\lambda)p(\lambda)d\lambda \\ &= \int_0^\infty \left(\frac{\lambda}{\beta}\right)^n \frac{s^{n-1}}{\Gamma(n)} \exp\left(-\frac{s\lambda}{\beta}\right) \lambda^{\alpha-1} \exp(-\beta\lambda) d\lambda \\ &= \frac{\beta^\alpha s^{n-1}}{\Gamma(n)\Gamma(\alpha)} \int \lambda^{n+\alpha-1} \exp(-\lambda(\beta+s)) \\ &= \frac{\beta^\alpha s^{n-1}}{\Gamma(n)\Gamma(\alpha)} \times \frac{\Gamma(n+\alpha)}{(\beta+s)^{n+\alpha}} . \end{aligned}$$

### ۴ تابع زیان پیشگیرانه<sup>۸</sup>

اولین بار نورستروم<sup>۹</sup> (۱۹۹۶) تابع زیان نامتقارن پیشگیرانه را معرفی کرد. وی در مقاله‌ی خود، تحلیل ریسک را در چهارچوب بیز مطرح می‌کند. در تحلیل ریسک، هم به توانایی وقوع یک پیشامد نامطلوب و هم به پیامدهای آن رسیدگی می‌شود.

سپس تیلمن<sup>۱۰</sup> (۲۰۰۳) انواعی از توابع زیان پیشگیرانه معرفی کرد، از جمله:

$$L(\theta, \delta) = \left[ \ln\left(\frac{\delta}{\theta}\right) \right]^2 \quad (۱.۴)$$

محاسبه‌ی برآورد بیز برای رابطه‌ی فوق، دشوار است. در واقع مسئله توسط یک تابع زیان مطرح می‌شود که هم محاسبات آسان داشته و هم بطور حسی خوشایند باشد.

برآوردگرهای محافظه‌کار زمانی که کم برآوردی ممکن است منجر به پیامد جدی شود، خیلی مفید هستند. یک تابع زیان پیشگیرانه‌ی

نامتقارن ساده و خیلی مفید، به صورت زیر می‌باشد:

$$L(\delta, \theta) = \frac{(\delta - \theta)^2}{\delta} \quad (۲.۴)$$

<sup>۷</sup>Bernardo

<sup>۸</sup>Precautionary

<sup>۹</sup>Norstrom

<sup>۱۰</sup>Tillman

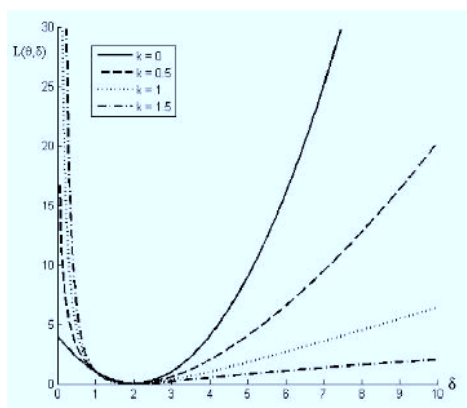
این تابع زیان بصورت زیر تعمیم داده می‌شود:

$$L(\theta, \delta) = \frac{(\delta - \theta)^2}{\delta^k} \omega(\theta) \quad ; \quad 0 < k \leq 2, \omega(\theta) > 0 \quad (3.4)$$

به طوری که  $\omega(\theta)$  یک تابع وزنی اختیاری است و  $k$  را شاخص پیشگیرانه می‌نامند.

تابع زیان (۳.۴)، وقتی  $k = 0$  باشد همان تابع زیان درجه دوم است و در صورتی که  $k = 2$ ، وقتی که  $\delta \rightarrow \infty$  به  $\omega(\theta)$  میل می‌کند. بنابراین، هنگامی که  $k$  افزایش یابد تابع زیان، بیشتر پیشگیرانه می‌شود.

توابع زیان (۳.۴) طیفی از توابع زیان پیشگیرانه را پوشش می‌دهد. در شکل زیر این تابع زیان برای  $k$ های مختلف رسم شده است.



شکل ۱: تابع زیان (۳.۴) برای های  $k$  مختلف.

## ۵ تعیین حجم نمونه به روش بیز

فرض کنیم  $p(\theta|x)$  تابع چگالی پسین باشد، پس ریسک پسین تحت این تابع زیان به صورت زیر خواهد بود:

$$PR = E\left(\frac{(\hat{\delta} - \theta)^2}{\hat{\delta}^k}\right)$$

اگر قرار دهیم

$$\frac{\partial PR}{\partial \hat{\delta}} = 0$$

در نتیجه

$$\hat{\delta} = \sqrt{E(\theta^2|x)}$$

که  $\hat{\delta}$  برآوردگر بیز  $\theta$  است. پس داریم:

$$\begin{aligned} PR &= E\left(\frac{(\hat{\delta} - \theta)^2}{\hat{\delta}^k}\right) \\ &= E\left(\frac{\hat{\delta}^2 + \theta^2 - 2\hat{\delta}\theta}{\hat{\delta}^k}\right) \\ &= 2(\hat{\delta} - E(\theta|x)) \\ &= 2(\sqrt{E(\theta^2|x)} - E(\theta|x)) \end{aligned}$$

با توجه به رابطه‌ی (۱۰۲) باید حجم نمونه را طوری بدست آوریم که مقدار تابع هزینه‌ی  $TC(n) = c_0 + cn + PR$  مینیمم شود. بنابراین مقداری از  $n$  که تابع زیر را مینیمم می‌کند را به عنوان حجم نمونه‌ی مطلوب در نظر می‌گیریم.

$$TC(n) = c_0 + cn + 2(\sqrt{E(\theta^2|x)} - E(\theta|x)) .$$

## ۶ تعیین حجم نمونه برای توزیع نمایی

فرض کنید  $X_1, X_2, \dots, X_n$  یک نمونه‌ی تصادفی به حجم  $n$  از توزیع نمایی با پارامتر  $\lambda$  باشد که چگالی احتمال آن به صورت زیر است:

$$p(x|\lambda) = \lambda \exp(-\lambda x) \quad ; \quad x > 0, \lambda > 0$$

فرض می‌کنیم  $S = \sum_{i=1}^n X_i$  پس تابع درستنمایی نمونه به صورت زیر است:

$$L(\lambda) = p(x|\lambda) = \lambda^n \exp(-\lambda s)$$

اگر توزیع پیشین را گاما با پارامترهای  $\alpha$  و  $\beta$  در نظر بگیریم، یعنی

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta \lambda)$$

در نتیجه چگالی پسین به صورت زیر بدست می‌آید:

$$p(\theta|x) \propto \lambda^{\alpha+n-1} \exp(-(\beta+s)\lambda)$$

یعنی توزیع پسین، گاما با پارامترهای  $\alpha+n$  و  $\beta+s$  می‌باشد

$$\theta|x \sim g(\alpha+n, \beta+s)$$

اگر  $\delta$  برآوردگر پارامتر  $\theta$  باشد، تحت تابع زیان پیشگیرانه برآورد بیز به صورت زیر خواهد بود:

$$\begin{aligned} \hat{\delta} &= \sqrt{E(\theta^2|x)} \\ &= \sqrt{E^2(\theta|x) + \text{var}(\theta|x)} \\ &= \sqrt{\left(\frac{\alpha+n}{\beta+s}\right)^2 + \frac{\alpha+n}{(\beta+s)^2}} \end{aligned}$$

پس تابع هزینه‌ی کل به صورت زیر می‌باشد:

$$\begin{aligned} TC(n) &= c_0 + cn + PR \\ &= c_0 + cn + 2\left(\sqrt{\left(\frac{\alpha+n}{\beta+s}\right)^2 + \frac{\alpha+n}{(\beta+s)^2}} - \frac{\alpha+n}{\beta+s}\right) . \end{aligned}$$

ملاحظه می‌شود تابع هزینه‌ی فوق از طریق  $S$  به مقادیر بردار تصادفی  $X_1, X_2, \dots, X_n$  بستگی دارد از این رو نسبت به  $S$  امید ریاضی می‌گیریم تا مستقل از مقادیر تصادفی بردار  $X$  شود.

در بخش ۳ نشان دادیم که  $S$  از توزیع گاما-گاما با پارامترهای  $\alpha$ ،  $\beta$  و  $n$  پیروی می‌کند یعنی

$$s \sim g_g(t|\alpha, \beta; n) ,$$

در نتیجه تابع چگالی  $S$  به صورت زیر می‌باشد:

$$p(s) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+n)}{\Gamma(n)} \frac{s^{n-1}}{(\beta+s)^{\alpha+n}} .$$

پس داریم:

$$E_*(TC(n)) = c_0 + cn + \gamma E_* \left( \sqrt{\left(\frac{\alpha+n}{\beta+s}\right)^2 + \frac{\alpha+n}{(\beta+s)^2}} - \frac{\alpha+n}{\beta+s} \right) ,$$

در نتیجه

$$E_*(TC(n)) = c_0 + cn + \gamma \left( \frac{\alpha \sqrt{(\alpha+n)(\alpha+n+1)}}{\beta(\alpha+n)} - \frac{\alpha}{\beta} \right) . \quad (1.6)$$

مقداری از  $n$  که رابطه‌ی (۱.۶) را مینیمم کند مقدار حجم نمونه‌ی ما خواهد بود که با استفاده از نرم افزارهای ریاضی می‌توان آن را بدست آورد.

به عنوان مثال با در نظر گرفتن  $\alpha = \beta = 1$  و  $\alpha = 0.5, \beta = 0.1$  برای هزینه‌های مختلف حجم نمونه را محاسبه می‌کنیم.

$$\alpha = \beta = 1 \Rightarrow \begin{cases} c = 0.001 \Rightarrow n \approx 31 \\ c = 0.01 \Rightarrow n \approx 9 \\ c = 0.02 \Rightarrow n \approx 6 \\ c = 0.03 \Rightarrow n \approx 5 \end{cases}$$

$$\alpha = 0.5, \beta = 0.1 \Rightarrow \begin{cases} c = 0.001 \Rightarrow n \approx 70 \\ c = 0.01 \Rightarrow n \approx 21 \\ c = 0.02 \Rightarrow n \approx 15 \\ c = 0.03 \Rightarrow n \approx 12 \end{cases}$$

## نتیجه‌گیری

در این مقاله نظریه‌ی تصمیم بیزی برای تعیین حجم نمونه تحت تابع زیان نامتقارن پیشگیرانه را به کار بردیم که این تابع زیان بسیار محتاطانه عمل می‌کند و از بیش و کم برآوردی جلوگیری می‌کند. از طرفی استفاده از روش بیز در نمونه‌گیری به این دلیل که بوسیله‌ی توزیع پیشین از قبل اطلاعات مفیدی داریم بسیار حائز اهمیت است. سرانجام پس از بدست آوردن حجم نمونه‌ی مطلوب، هزینه‌ی کلی نمونه‌گیری را مینیمم کردیم.

## مراجع

Bernardo, J. M., Smith, A. F. M. (1994). Bayesian Theory. Chichester: John Wiley.

Grundy, P. M., Healy, M. J. R. and Rees, D. H. (1956). Economic choice of the amount of experimentation.

J. R. Statist. Soc. A 18, 32-48.

- Lindley, D. V. East, D. A. and Hamilton, P. A. (1960), Tables for making Inferences about the Variance of a Normal Distribution. *Biometrika*, 47: 433-437.
- Lindley, D. V. (1997b). The choice of sample size – A reply to the discussion. *Statistician* 46:163–166.
- Norstrom, J. G. (1996), The use of precautionary loss function in risk analysis. *IEEE Trans. Reliab.*, Vol. 45(3), pp. 400-403.
- O'Hagan, A., Stevens, J. W., (2001). Bayesian assessment of Sample size for clinical trails of cost-effectiveness. *Med. Decision Making*, 21, 219-230 .
- Pezeshk, H., Gittins, J. C., (1999). Sample size Determination in Clinical Trials. *Student*3, 19-26.
- Pezeshk, H., (2003). Bayesian techniques for Sample size determination in clinical trials: A short review. *Statistical Method in Medical Research*, 12 (6), 489-504 .
- Stallard, N., (1998). Sample size determination for phase II clinical trials based on Bayesian decision theory. *Biometrics*, 54,279-294.
- Tillman, F. A., Kuo, W., Hwang, C. L. and Doris, L.G. (1982), Bayesian reliability availability A review. *IEEE Trans. Reliab.*, vol R-31, pp. 362-372.