# آشنایی با نظریه انحرافات بزرگ

حسین محمدی، صفیه محمودی
گروه آمار، دانشکده علوم ریاضی، دانشگاه صنعتی اصفهان

**چکیده**

قضایای حدی یکی از مهم‌ترین نتایج نظری در نظریه احتمال هستند که از جمله قضیه‌های مهم آن قضیه حد مرکزی و قانون اعداد بزرگ می‌باشد. به بیان ساده، قضیه حد مرکزی بیان می‌کند که مجموع تعداد زیادی از متغیرهای تصادفی مستقل و هم‌توزیع به سمت یک متغیر تصادفی مشخص میل می‌کند که تقریبا دارای توزیع نرمال است و در قانون اعداد بزرگ شرایطی بیان می‌شود که تحت آن شرایط میانگین دنباله‌ای از متغیرهای تصادفی به متوسط امید ریاضی خود همگرا باشند. قضیه حد مرکزی نسبت به قانون اعداد بزرگ اطلاعات بیشتری در مورد رفتار میانگین دنباله ارائه می‌دهد اما در مورد نحوه همگرایی توزیع دنباله چیزی بیان نمی‌کند. در این مقاله به نظریه انحرافات بزرگ پرداخته می‌شود که در آن به نحوه همگرایی و یا سرعت همگرایی چنین دنباله‌هایی می‌پردازد.

**واژه‌های کلیدی:** انحرافات بزرگ، قضیه حد مرکزی، قانون قوی اعداد بزرگ.

## ۱ مقدمه

مفهوم نظریه انحرافات بزرگ[۱] اولین بار در سال ۱۹۳۲ توسط اسشر[۲] [۸] که علاقه‌مند به تجزیه و تحلیل احتمال خطر در صنعت بیمه بود، به کار برده شد. همچنین نتایج انحرافات بزرگ برای مجموع متغیرهای تصادفی مستقل توسط کرامر[۳] [۲] در سال ۱۹۳۸، و برای متغیرهای تصادفی مستقل و هم‌توزیع توسط سانو[۴] [۱۴] در سال ۱۹۵۷ بیان شد. البته این نظریه به شیوه‌ای که امروزه می‌شناسیم توسط وارادان[۵] [۱۶] در سال ۱۹۶۶ معرفی شد.

تعمیم نتایج نظریه انحرافات بزرگ به فرآیندها و

[۱] Large Deviations Theory
[۲] Esscher
[۳] Cramer
[۴] Sanov
[۵] Varadhan

زنجیرهای مارکف در مقالات متعددی، از جمله دونسکر و وارادان[۶] [۴، ۵، ۶] و گارتنر[۷] [۱۰] بیان شده است. همچنین وارادان [۱۵] در سال ۲۰۰۳ در مبحث قدم زدن تصادفی نظریه انحرافات بزرگ را وارد کرد. در سال ۲۰۱۲ بروکو و موگولسکی[۸] [۱] هم در این زمینه به تحقیق پرداختند. البته در زمینه انحرافات بزرگ افراد دیگری همچون الیس[۹] [۷]، دمبو و زیتونی[۱۰] [۳]، فنگ و کورتز[۱۱] [۹] نیز تحقیقات مهمی انجام داده‌اند.

## ۲ مقدماتی از انحرافات بزرگ

در ابتدا با بیان چند مثال ساده به تشریح مفهوم انحرافات بزرگ پرداخته می‌شود.

***مثال ۱.۲.*** *اگر سکه سالمی $n$ بار پرتاب شود، در هر پرتاب سکه دو حالت امکان پذیر است و تعداد کل حالات در $n$ پرتاب برابر با $2^n$ است. همچنین تعداد شیرها $n+1$ مقدار (از ۰ تا $n$) ممکن است باشند. اگر $\overline{X}_n$ میانگین تعداد شیرها در $n$ پرتاب باشد، واضح است که $\overline{X}_n$ یکی از مقادیر $0, \frac{1}{n}, \ldots, 1$ را اختیار می‌کند.*

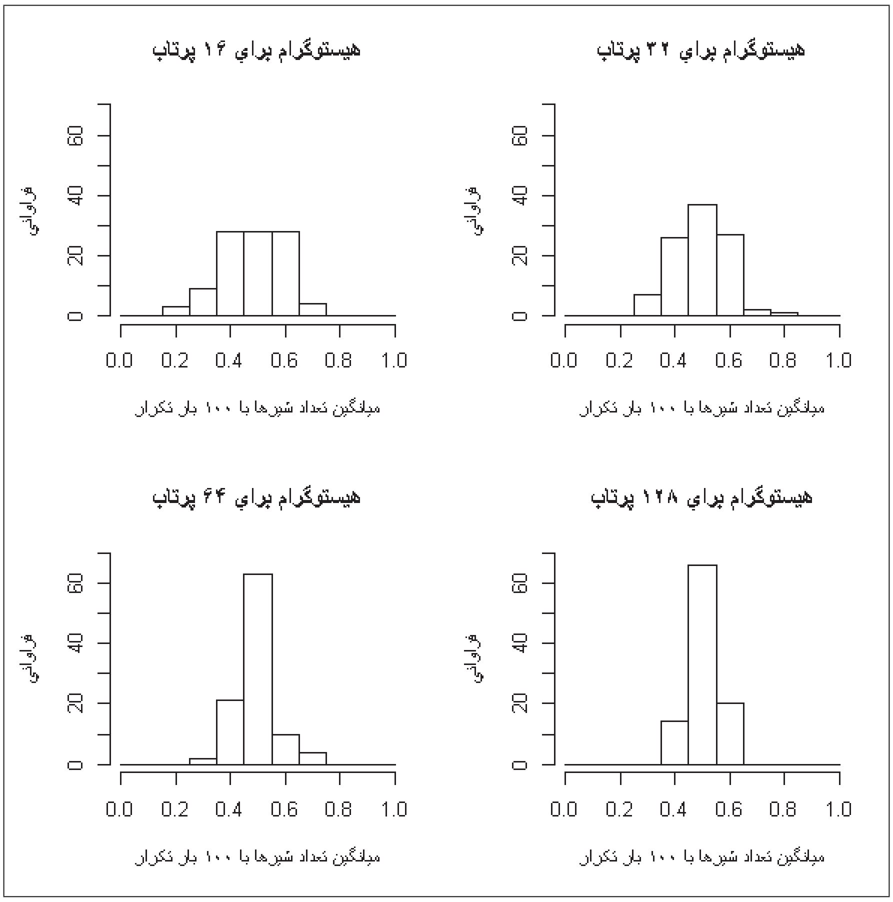
*در این مثال ابتدا هیستوگرام $\overline{X}_n$ برای مقادیر $n = 16$، $n = 32$، $n = 64$ و $n = 128$ رسم می‌شود. مثلاً برای رسم هیستوگرام $\overline{X}_n$ در حالت $n = 16$، فرض می‌شود سکه سالمی ۱۶ بار پرتاب شده باشد و میانگین تعداد شیرها یادداشت شده باشد و این آزمایش ۱۰۰ بار تکرار شود و هیستوگرام میانگین تعداد شیرها برای این ۱۰۰ تکرار رسم شود. به همین ترتیب این آزمایش برای $n = 32$، $n = 64$ و $n = 128$ تکرار می‌شود. همان‌طور که در شکل ۱ دیده می‌شود با افزایش $n$، $\overline{X}_n$ به مقدار $\frac{1}{2}$ (میانگین جامعه) نزدیکتر می‌شود، یا به عبارتی با افزایش $n$، دم‌های دنباله کوچک و کوچکتر می‌شوند.*

*به عبارت دقیق‌تر طبق قانون اعداد بزرگ میانگین نتایج بدست آمده از آزمایش‌ها با احتمال یک به مقدار میانگین جامعه همگرا می‌شود. بنابراین احتمال پیشامدهایی که شامل میانگین نباشند، به صفر میل می‌کند. آنچه در نظریه انحرافات بزرگ مورد توجه است سرعت این همگرایی است و این‌که دم‌های دنباله با چه نرخی به سمت صفر میل می‌کنند. اصل انحرافات بزرگ اصولاً به بررسی پدیده‌هایی با احتمال ضعیف و یا به اصطلاح پیشامدهای نادر می‌پردازد.*

*با فرض $a = 0/6$، همان‌طور که با توجه به شکل ۲ دیده می‌شود با افزایش $n$، $\log P(\overline{X}_n > 0/6)$ تقریباً به صورت خطی در حال کاهش است. زیرا با توجه به شکل ۱ با افزایش $n$، $P(\overline{X}_n > a)$ به سمت صفر میل می‌کند و بنابراین لگاریتم آن به سمت $-\infty$ می‌رود. در شکل ۳، $\log P(\overline{X}_n > a)$ برای مقادیر $a = 0/6$، $a = 0/7$، $a = 0/8$ و $a = 0/9$ رسم شده است. با توجه به شکل ۳ دیده می‌شود که هر چقدر $a$ از $\frac{1}{2}$ (میانگین جامعه) دورتر شود $\log P(\overline{X}_n > a)$ با سرعت بیشتری نزول می‌کند.*

*در شکل ۲ و ۳ مقدار $\log P(\overline{X}_n > a)$ با فرمول دقیق محاسبه شده است و شبیه‌سازی نمی‌باشد. با توجه به توضیحات داده شده، دیده می‌شود که هر چه $\log P(\overline{X}_n > a)$ با شیب بیشتری نزول کند، میانگین نمونه سریع‌تر به میانگین جامعه همگرا می‌شود. به طور مشابه می‌توان نشان داد که اگر $a$ کمتر از $\frac{1}{2}$ باشد، تمام مطالب گفته شده*

[۶] Donsker and Varadhan
[۷] Gärtner
[۸] Borovkov and Mogulskii
[۹] Ellis
[۱۰] Dembo and Zeitouni
[۱۱] Feng and Kurtz




شکل ۱: هیستوگرام میانگین تعداد شیرها برای $n = 16$، $n = 32$، $n = 64$ و $n = 128$

*برای* $\log P(\overline{X}_n < a)$ *نیز به صورت معادل برقرار است.*

حال در حالت کلی‌تر فرض می‌شود $X_1$،$X_2$،...، دنباله‌ای از متغیرهای تصادفی مستقل، هم‌توزیع و حقیقی‌مقدار با میانگین متناهی $m$ باشند، آن‌گاه طبق قانون قوی اعداد بزرگ، وقتی $n \to \infty$، با احتمال یک،

$$\frac{1}{n}(X_1 + \cdots + X_n) \to m.$$

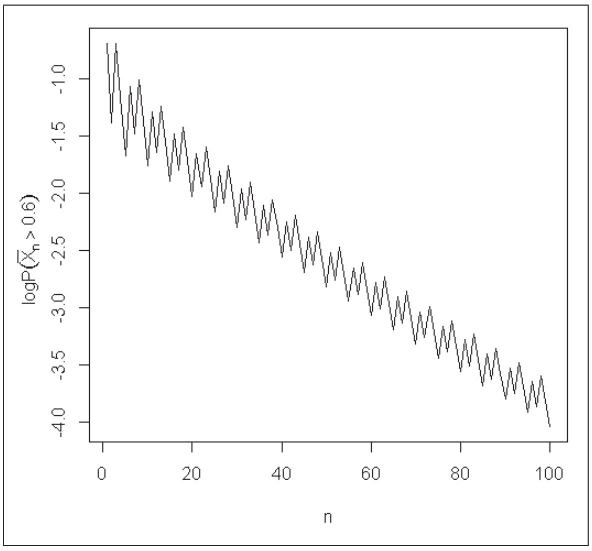
بنابراین اگر $a > m$، آن‌گاه،

$$\lim_{n\to\infty} P\left(\frac{S_n}{n} > a\right) = 0,$$

که $S_n = X_1 + \cdots + X_n$.
آنچه از قضیه حد مرکزی در مورد نرخ همگرایی حد فوق بدست می‌آید، وقتی $n \to \infty$ به صورت زیر است :

$$\begin{aligned} P\left(\frac{S_n}{n} > a\right) &= P\left(\frac{S_n - nm}{\sqrt{n}} > \sqrt{n}\,(a-m)\right) \\ &\approx 1 - \Phi\left(\sqrt{n}\,(a-m)\right) \to 0, \quad (1) \end{aligned}$$

که برای هر $x \in R$، $\Phi(x)$ تابع توزیع نرمال است. همان‌طور که از تقریب بالا معلوم است، قضیه حدمرکزی تقریب واضحی از نرخ همگرایی نمی‌دهد و فقط بیان می‌کند هر چقدر $a$ از $m$ بزرگتر باشد دنباله سریع‌تر به صفر همگراست.
شبیه‌سازی رابطه (۱) در شکل ۴ آمده است که در آن $X_i$ها



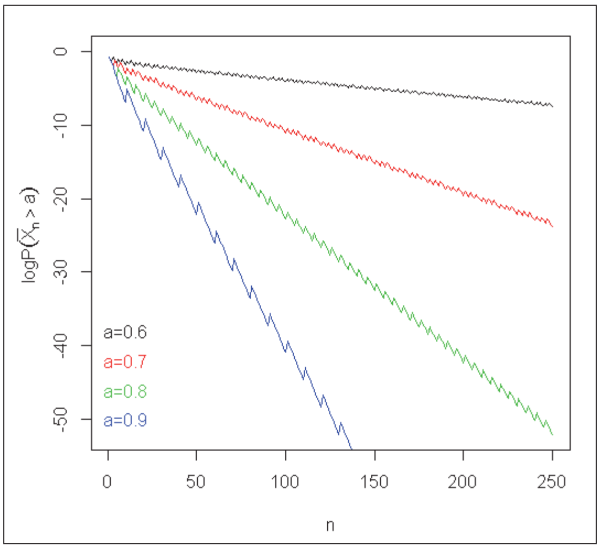

شکل ۲: نمودار $\log P(\overline{X}_n > ۰/۶)$ در مقابل $n$

متغیرهای تصادفی مستقل و هم‌توزیع، با توزیع نمایی با پارامتر ۲ فرض شده است $(m = \frac{۱}{۲})$.

هر چند که پیشامد $\{\frac{S_n}{n} > a\}$ یک پیشامد نادر است، اما این پیشامد در مسائل بیمه قابل توجه و رایج است. به طور مثال اگر $X_i$ ادعای خسارت مشتری $i$ام از شرکت بیمه در یک سال باشد، هر چند که سرمایه اولیه $(na)$ شرکت‌های بیمه خیلی زیاد است، ولی احتمال $\frac{S_n}{n} > a$ مهم است و برابر احتمال ورشکستگی شرکت بیمه است. بنابراین عجیب نیست که بررسی پیشامدهای نادر مورد توجه محققان علم بیمه، به ویژه احتمالدان سوئدی کرامر که یک پیشرو در مدل‌سازی ریاضی از مسائل بیمه و آمار بود قرار گرفته باشد [۲].

***تعریف ۱.۲.*** *اگر $\Omega$ یک فضای توپولوژی باشد، تابع $I : \Omega \to R \cup \{-\infty, \infty\}$ نیمه‌پیوسته پایین[۱۲] نامیده می‌شود، هرگاه،*

$$\liminf_{x \to x_۰} I(x) \geq I(x_۰).$$




شکل ۳: نمودار $\log P(\overline{X}_n > a)$ در مقابل $n$ برای $a = ۰/۶$، $a = ۰/۷$، $a = ۰/۸$ و $a = ۰/۹$

***تعریف ۲.۲.*** *اگر $\Omega$ یک فضای توپولوژی باشد، آن‌گاه تابع $I : \Omega \to [۰, \infty]$ را یک تابع نرخ[۱۳] می‌نامند، هرگاه یک نگاشت نیمه‌پیوسته پایین باشد، به‌طوری‌که برای هر $\alpha \in [۰, \infty)$ مجموعه $\Psi_I(\alpha) = \{x : I(x) \leq \alpha\}$ یک زیرمجموعه بسته از $\Omega$ باشد. اگر تمام مجموعه‌ها به فرم $\Psi_I(\alpha)$ زیرمجموعه‌های فشرده از $\Omega$ باشند، آن‌گاه تابع نرخ، تابع نرخ خوب[۱۴] نامیده می‌شود [۳].*

***مثال ۲.۲.*** *اگر $\{X_i\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع، با توزیع $N(۰, ۱)$ باشند، آن‌گاه میانگین نمونه دارای توزیع $N\left(۰, \frac{۱}{n}\right)$ خواهد بود. از طرفی برای هر $a > ۰$ می‌توان نشان داد که:*

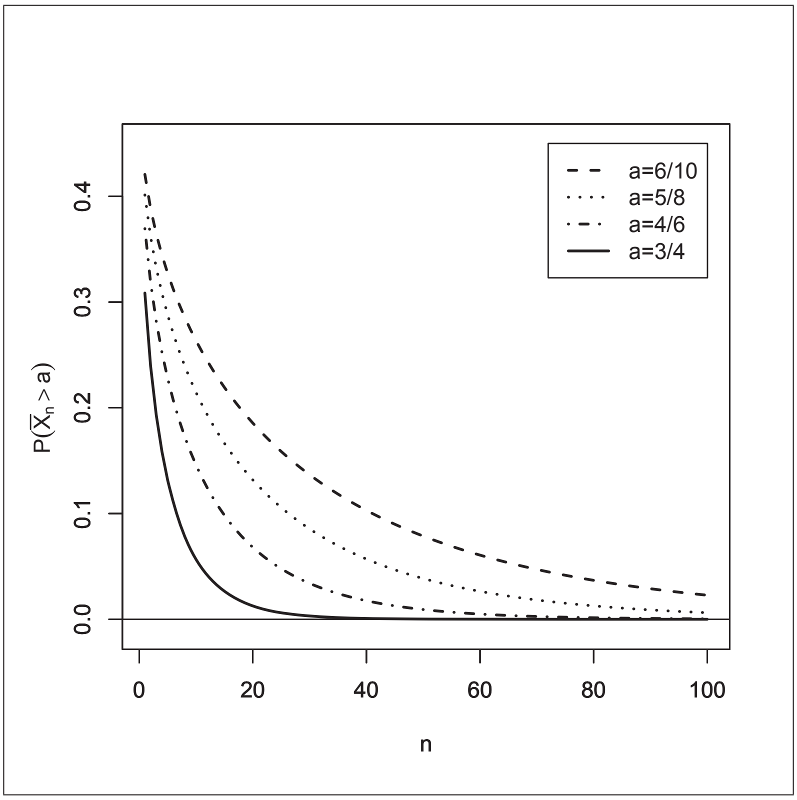
$$\lim_{n \to \infty} \frac{۱}{n} \log P\left(|\frac{۱}{n} S_n| \geq a\right) \approx \frac{-a^۲}{۲}.$$

*از رابطه بالا نتیجه می‌شود که احتمال پیشامد نادر $\{|\frac{۱}{n} S_n| \geq a\}$ از مرتبه $\exp\left(\frac{-۱}{۲} n a^۲\right)$ است. این یک*

[۱۲] Lower Semi-Continuity

[۱۳] Rate Function

[۱۴] Good Rate Function




شکل ۴: شبیه‌سازی رابطه (۱)

*مثال معمول در نظریه انحرافات بزرگ است که در آن $I(a)=\frac{1}{2}a^2$ را تابع نرخ می‌نامند [۱۲]. شبیه‌سازی این مثال به صورت زیر است.*

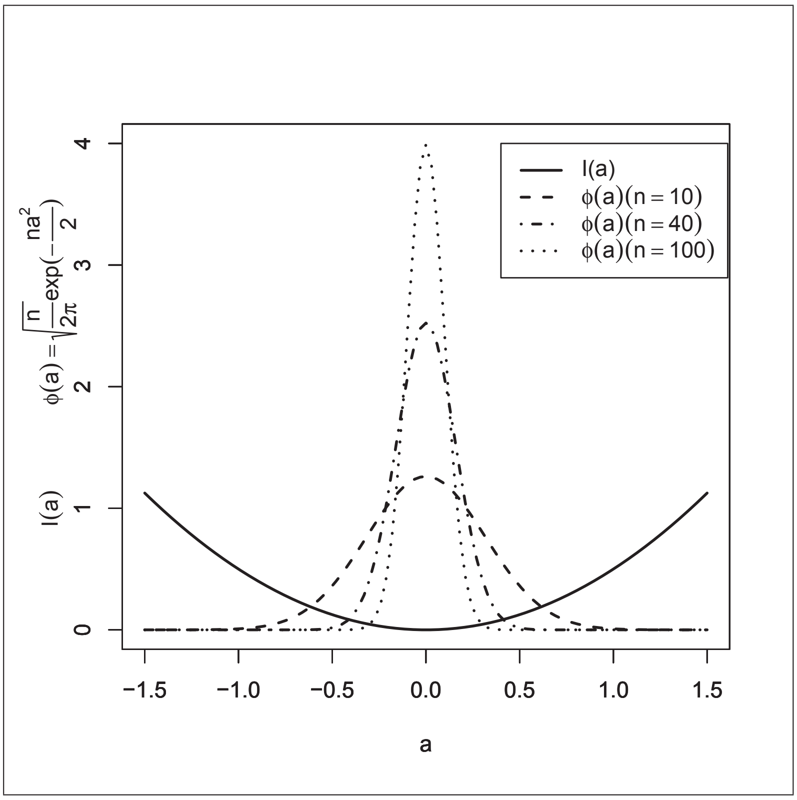
*همانطور که بیان شد، $\frac{1}{n}S_n \sim N(0,\frac{1}{n})$ است در شکل ۵ تابع $I(a)=\frac{1}{2}a^2$ و $\phi(a)$ (تابع چگالی میانگین نمونه) به ازای $n=10$، $n=40$ و $n=100$ رسم شده است.*

*با توجه به شکل ۵ دیده می‌شود که با افزایش $n$ دم‌های تابع چگالی $\frac{1}{n}S_n$ به سمت صفر میل می‌کند و تعبیر آن همانند مثال ۱.۲ می‌باشد. به عبارت دیگر با افزایش $n$، واریانس میانگین نمونه در حال کاهش است. همچنین در شکل ۵ تابع $I(a)$ رسم شده است که نشان می‌دهد هر چه $a$ به مقدار میانگین جامعه نزدیکتر باشد نرخ همگرایی کمتر است.*

کرامر اولین دانشمندی بود که در نظریه انحرافات بزرگ به بیان قضیه پرداخت. قضیه کرامر تعمیمی از مثال‌های فوق برای یک دنباله دلخواه از متغیرهای مستقل و هم‌توزیع است و به عنوان اولین قضیه در نظریه انحرافات بزرگ شناخته می‌شود.

**قضیه ۱.۲.** *اگر $X_1$، $X_2$، ...، متغیرهای تصادفی مستقل و هم‌توزیع با تابع چگالی $f$ و تابع مولد گشتاور متناهی $M$ باشند. یعنی برای هر $t \in R$،*

$$\begin{aligned} M(t) &:= E\left(e^{tX}\right) \\ &= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx < \infty, \end{aligned}$$




شکل ۵: تابع چگالی $\phi(a)$ برای $n=$ ۱۰، $n=$ ۴۰ و $n=$ ۱۰۰، و تابع $I(a)$ در مقابل $a$

*آن‌گاه برای هر* $a > E(X_1)$،

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}S_n \geq a\right) = -I\left(a\right),$$

*که در آن*

$$I(a) = \sup_{t\in R}\{at - \log\left(M\left(t\right)\right)\}.$$

*و به طور مشابه برای* $a < E\left(X_1\right)$،

$$\lim_{n\to\infty}\frac{1}{n}\log P\left(\frac{1}{n}S_n \leq a\right) = -I\left(a\right).$$

*اثبات.* برای اثبات می‌توان به [۱۱] مراجعه کرد. □

براساس قضیه حد مرکزی معلوم است که $\overline{X}_n$ دارای توزیع نرمال مجانبی حول $E[X_1]$، به ازای مقادیر بزرگ $n$ است. همچنین قانون قوی اعداد بزرگ بیان می‌کند که هر چقدر مقدار $n$ بزرگتر شود، با احتمال ۱، $\overline{X}_n$ به $E[X_1]$ نزدیکتر خواهد شد. قضیه حد مرکزی نسبت به قضیه اعداد بزرگ اطلاعات بسیاری را می‌تواند در مورد رفتار $S_n$ ارائه دهد. مثلا با استفاده از قضیه حد مرکزی می‌توان توزیع مجانبی $\overline{X}_n$ یا $P(\overline{X}_n > x)$ را محاسبه کرد. در واقع قضیه حد مرکزی هر چند که در مورد نحوه همگرایی توزیع دنباله وقتی $n\to\infty$ چیزی بیان نمی‌کند، اما اطلاعاتی در مورد نحوه توزیع داده‌ها در نزدیکی نقطه حدی در اختیار می‌گذارد. قضیه ۱.۲ به نحوهٔ همگرایی و یا سرعت همگرایی چنین احتمالاتی می‌پردازد، که در بسیاری از شرایط در عمل از اهمیّت به‌سزایی برخوردار

است. به همین دلیل نیز در مسائل کاربردی از این روش استفاده زیادی شده است.

## ۳ اصل انحرافات بزرگ

بعد از پرداختن به مقدمات، در این بخش به بیان تعریف اصل انحرافات بزرگ و قضایای مربوط به آن پرداخته می‌شود.

**تعریف ۱.۳.** *خانواده $\{Z_n, n > 0\}$ روی فضای توپولوژی $\Omega$، در اصل انحرافات بزرگ با تابع نرخ $I$ و سرعت $n$ صدق می‌کند، اگر تابع $I : \Omega \to [0, \infty]$ نیمه‌پیوسته پایین باشد و برای هر مجموعه بسته $C$ نامساوی،*

$$\limsup_{n\to\infty} \frac{1}{n} \log P(Z_n \in C) \leq -\inf_{a\in C} I(a),$$

*و همچنین برای هر مجموعه باز $D$ نامساوی،*

$$\liminf_{n\to\infty} \frac{1}{n} \log P(Z_n \in D) \geq -\inf_{a\in D} I(a),$$

*برقرار باشد.*

**مثال ۱.۳.** *اگر $\Omega = [0, 1]$ باشد آن‌گاه دنباله پرتاب متوالی سکه سالم‌که در مثال ۱.۲ مطرح گردید، در اصل انحرافات بزرگ با تابع نرخزیر صدق می‌کند.*

$$I(a) = a \log a + (1 - a) \log(1 - a) + \log 2$$

*برای این منظور با توجه به تعریف ۱.۳ کافی‌است برای هر مجموعه باز $D$،*

$$\liminf_{n\to\infty} \frac{1}{n} \log P(\overline{X}_n \in D) \geq -\inf_{a\in D} I(a), \qquad (2)$$

*و برای هر مجموعه بسته $C$،*

$$\limsup_{n\to\infty} \frac{1}{n} \log P(\overline{X}_n \in C) \leq -\inf_{a\in C} I(a), \qquad (3)$$

*اثبات شود.*

*اگر $D$ یک مجموعه باز باشد و $a \in D \cap [0, 1]$، آن‌گاه برای $n$های به قدرکافی بزرگ $\frac{\lfloor na \rfloor}{n} \in D$ . بنابراین،*

$$P(\overline{X}_n \in D) \geq P\left(\overline{X}_n = \frac{\lfloor na \rfloor}{n}\right).$$

*از طرفی با استفاده از فرمول استرلینگ،*

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\overline{X}_n = \frac{\lfloor na \rfloor}{n}\right)$$
$$= -a \log \frac{a}{p} - (1 - a) \log \frac{1 - a}{1 - p}.$$

*بنابراین با توجه به این‌که $p = 0/5$ برای هر $a \in D \cap [0, 1]$، نامساوی زیر برقرار است:*

$$\lim_{n\to\infty} \frac{1}{n} \log P(\overline{X}_n \in D)$$
$$\geq -a \log a - (1 - a) \log(1 - a) - \log 2.$$

*لذا رابطه (۲) اثبات می‌شود به همین ترتیب برای هر مجموعه بسته $C$، می‌توان درستی رابطه (۳) را نشان داد [۱۳].*

در پایان، قضیه گارتنر-الیس[۱۵]، که تعمیمی از قضیه کرامر است بیان می‌شود.

**قضیه ۱.۳.** *اگر $\{X_n\}$ یک خانواده از متغیرهای تصادفی باشد، که در $R^d$ مقدار می‌گیرند، به طوری‌که،*

$$\Lambda(a) := \lim_{n\to\infty} \frac{1}{n} \log E\left(e^{n\langle a, X_n\rangle}\right),$$

[۱۵] Gärtner-Ellis

*برای هر* $a \in R^d$ *موجود و متناهی باشد که* $\langle a, X_n\rangle$ *ضرب داخلی اقلیدسی است. اگر* $\Lambda$ *مشتق‌پذیر باشد، آن‌گاه متغیرهای تصادفی* $X_n$ *در اصل انحراف بزرگ، با سرعت* $n$ *و تابع نرخ خوب زیر صدق می‌کنند.*

$$I(a) := \sup_{t\in R^d}\{\langle a,t\rangle - \Lambda(t)\}.$$

*اثبات.* برای اثبات می‌توان به قضیه ۲.۳.۶ مرجع [۳] و یا قضیه ۱۲.۱۴ مرجع [۱۳] مراجعه کرد. $\square$

# مراجع


[1] Borovkov, A. A. and Mogulskij, A. A. (2012). Chebyshev-type exponential inequalities for sums of random vectors and for trajectories of Rrandom walks. Theory Probab. Appl., 56(1), 21-43.

[2] Cramer, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. Colloque Consecréàla Théorie des Probabilités, 736, 5-23.

[3] Dembo, A. and Zeitouni, O. (2010). Large deviations techniques and applications. Spinger, Berlin.

[4] Donsker, M. D. and Varadhan, S. R. S. (1975). Asymptotic evaluation of certain markov process expectations for large time. Communications on Pure and Applied Mathematics, 28, 1-47.

[5] Donsker, M. D. and Varadhan, S. R. S. (1976). Asymptotic evaluation of certain markov process expectations for large time. Communications on Pure and Applied Mathematics, 29, 389-461.

[6] Donsker, M. D. and Varadhan, S. R. S. (1983). Asymptotic evaluation of certain markov process expectations for large time. Communications on Pure and Applied Mathematics, 36, 183-212.

[7] Ellis, R. S. (2006). Entropy, large deviations and statistical mechanics. Springer, Berlin.

[8] Esscher, F. (1932). On the probability function in the collective theory of risk. Skandinavisk Aktuarietidskrift, 15, 175-195.

[9] Feng, J. and Kurtz, T. G. (2006). Large deviations for stochastic processes, Mathematical Surveys and Monographs, 131.

[10] Gartner, J. (1977). On large deviations from an invariant measure.

Theory of Probability and its Applications, 22(1), 24-39. (Original in Russian edition: Theoria Veroyatnostei i ee Premaneniya, 22(1), 27-42).

[11] Ramasubramanian, S. (1999). Some aspects of large deviation. Stat. Math. Unit, Indian Statistical Institute.

[12] Ramasubramanian, S. (2008). Large deviations: an introduction to 2007 Abel Prize. Stat. Math. Unit, Indian Statistical Institute, 161-182.

[13] Rassoul-Agha, F. and Seppalainen, T. (2015). A course on large deviations with an introduction to gibbs measures, Graduate Studies in Mathematics, 162.

[14] Sanov, I. N. (1957). On the probability of large deviations of random magnitudes. Mat. Sb. N. S., 42, 11-44.

[15] Vardhan, S. R. S. (2003a). Large deviations and entropy. In Entropy. Princeton Ser. Appl. Math., 22, 199-214.

[16] Vardhan, S. R. S. (1966). Asymptotic probability and differential equations. Comm. Pure Appl. Math., 19, 261-286.