فصل ۷: دنبالهٔ متغیرهای تصادفی

Chapter 7

۱. مفاهیم کلی (pmf، CDF و pdf مشترک، استقلال، امید ریاضی، کواریانس، تابع مولد گشتاور، توابعی از n متغیر تصادفی، متغیرهای تصادفی مشترکاً نرمال، توزیع‌های شرطی)

۲. کاربردها (مجموع‌های تصادفی، نمونه‌برداری، آماره‌های رتبه، مجموع متغیرهای تصادفی مشترکاً نرمال)

۳. قضیهٔ حد مرکزی

۴. همگرایی دنبالهٔ متغیرهای تصادفی

۵. قانون اعداد بزرگ

۶. توابع توزیع متداول در آمار

در فصل ۵ دیدیم که چگونه مفاهیمی را که داشتیم به دو متغیر تصادفی توسعه دهیم. به طریق مشابه، وقتی چندتا متغیر تصادفی داشته باشیم، می‌توانیم مفاهیم گذشته (تابع CDF مشترک، pdf مشترک، مشروط و ...) را مطرح کنیم.

در ابتدا بردار $\underline{\mathbf{x}}$ را به صورت زیر تعریف می‌کنیم:

$$\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$$

pmf مشترک:

$$P_{\underline{\mathbf{x}}}(\underline{x}) = P_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \ldots, \mathbf{x}_n = x_n\}$$

که داریم:

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

CDF مشترک:

$$F_{\underline{\mathbf{x}}}(\underline{x}) = F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = P\{\mathbf{x}_1 \le x_1, \mathbf{x}_2 \le x_2, \ldots, \mathbf{x}_n \le x_n\}$$

مقدار این تابع همواره بین صفر و یک است. به ازای تمام آرگومان‌ها صعودی است. در نقطهٔ $(-\infty, -\infty, \cdots, -\infty)$ برابر صفر بوده و در نقطهٔ $(+\infty, +\infty, \cdots, +\infty)$ برابر یک می‌شود. اگر به جای برخی از آرگومان‌ها بی‌نهایت بگذاریم، CDF مشترک سایر متغیرهای تصادفی حاصل می‌شود. مثلاً داریم:

$$F_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(+\infty, x_2, x_3) = F_{\mathbf{x}_2\mathbf{x}_3}(x_2, x_3)$$

pdf مشترک:

$$f_{\underline{\mathbf{x}}}(\underline{x}) = f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\underline{\mathbf{x}}}(\underline{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n)$$

به این ترتیب روابط زیر برقرار خواهند بود:

$$f_{\underline{\mathbf{x}}}(\underline{x}) dx_1 dx_2 \cdots dx_n = P\{x_1 \le \mathbf{x}_1 \le x_1 + dx_1, x_2 \le \mathbf{x}_2 \le x_2 + dx_2, \ldots, x_n \le \mathbf{x}_n \le x_n + dx_n\}$$

$$F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(u_1, u_2, \ldots, u_n) du_n \cdots du_2 \, du_1$$

مقدار تابع pdf همواره مثبت است. انتگرال nگانهٔ آن از $-\infty$ تا $+\infty$ یک می‌شود. همچنین داریم:

$$\forall D \subset \mathbb{R}^n : P\{(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \in D\} = \int_D \cdots \int f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) dx_1 \cdots dx_n = \int_D f_{\underline{\mathbf{x}}}(\underline{x}) d\underline{x}$$

اگر از $-\infty$ تا $+\infty$ روی برخی از آرگومان‌ها انتگرال بگیریم، pdf مشترک سایر متغیرهای تصادفی حاصل می‌شود. مثلاً داریم:

$$f_{\mathbf{x}_2\mathbf{x}_3}(x_2, x_3) = \int_{-\infty}^{+\infty} f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(x_1, x_2, x_3) dx_1$$

استقلال:

متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ را مستقل گویند، هرگاه واقعه‌های $\{\mathbf{x}_1 \le x_1\}$، $\{\mathbf{x}_2 \le x_2\}$، ... و $\{\mathbf{x}_n \le x_n\}$ برای هر $\mathbf{x}_1$، هر $\mathbf{x}_2$، ... و هر $\mathbf{x}_n$ مستقل باشند.

معادلاً $\mathbf{x}_i$ها را مستقل گویند، هرگاه برای هر x_1، هر x_2، ... و هر x_n داشته باشیم:

$$F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = F_{\mathbf{x}_1}(x_1)F_{\mathbf{x}_2}(x_2)\cdots F_{\mathbf{x}_n}(x_n) = \prod_{i=1}^{n} F_{\mathbf{x}_i}(x_i)$$

یا معادلاً:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = f_{\mathbf{x}_1}(x_1)f_{\mathbf{x}_2}(x_2)\cdots f_{\mathbf{x}_n}(x_n) = \prod_{i=1}^{n} f_{\mathbf{x}_i}(x_i)$$

(در این صورت در مورد هر زیرمجموعه از اندیس‌ها نیز برقرار خواهد بود.)

مشابه حالت دو بعدی می‌توان ثابت کرد که اگر $\mathbf{x}_i$ها مستقل باشند، $g_i(\mathbf{x}_i)$ها هم مستقل خواهند بود.

همچنین داریم:

$$\underbrace{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m}_{\boxed{\text{گروه } A}},\underbrace{\mathbf{x}_{m+1},\ldots,\mathbf{x}_n}_{\boxed{\text{گروه } B}}$$

گروه A از گروه B مستقل است، اگر:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)=f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_m}(x_1,x_2,\ldots,x_m)f_{\mathbf{x}_{m+1}\cdots\mathbf{x}_n}(x_{m+1},\ldots,x_n)$$

امید ریاضی:

اگر $\mathbf{z} = g(\underline{\mathbf{x}})$ باشد، داریم:

$$E(\mathbf{z}) = \int_{-\infty}^{+\infty} z\, f_{\mathbf{z}}(z)\, dz$$

ولی بدون محاسبهٔ $f_{\mathbf{z}}$ نیز می‌توان $E(\mathbf{z})$ را حساب کرد:

$$E(\mathbf{z}) = E[g(\underline{\mathbf{x}})] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(\underline{x})\, f_{\underline{\mathbf{x}}}(\underline{x})\, d\underline{x}$$ (این انتگرال n بُعدی است)

و مشابه حالت دوبُعدی، از اینجا می‌توان خطی بودن امید ریاضی را ثابت کرد:

$$E\left[\sum_{i=1}^{n} a_i\, g_i(\underline{\mathbf{x}})\right] = \sum_{i=1}^{n} a_i\, E[g_i(\underline{\mathbf{x}})]$$ (یعنی جای امید و مجموع را می‌توان عوض کرد)

از جمله داریم:

$$E\left[\sum_{i=1}^{n} a_i \mathbf{x}_i\right] = \sum_{i=1}^{n} a_i E(\mathbf{x}_i)$$

برای درک نحوهٔ توزیع هر $\mathbf{x}_i$، η_i و σ_i به ما کمک می‌کنند که ایده‌ای کلی از نحوهٔ توزیع آن به دست آوریم. داریم:

$$\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \quad \rightarrow \quad \eta_{\underline{\mathbf{x}}} = E(\underline{\mathbf{x}}) = \begin{bmatrix} E(\mathbf{x}_1) \\ E(\mathbf{x}_2) \\ \vdots \\ E(\mathbf{x}_n) \end{bmatrix} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}$$ (بردار میانگین)

و نیز دیده بودیم که کواریانس دو متغیر تصادفی، نحوهٔ بستگی آنها را به هم نشان می‌داد. حال با کواریانس دو به دوی این n تا متغیر تصادفی رو به رو هستیم:

$$\sigma_{ij} \triangleq \mu_{\mathbf{x}_i\mathbf{x}_j} = \operatorname{cov}(\mathbf{x}_i, \mathbf{x}_j) = E\left[(\mathbf{x}_i - \eta_i)(\mathbf{x}_j - \eta_j) \right] : i, j : 1, 2, \ldots, n$$

$$\sigma_{ii} = \operatorname{cov}(\mathbf{x}_i, \mathbf{x}_i) = \operatorname{var}(\mathbf{x}_i) = \sigma_i^2$$

ماتریس کواریانس $\underline{\mathbf{x}}$ (Covariance Matrix) طبق تعریف برابر است با:

$$C_{\underline{\mathbf{x}}} = E\left[(\underline{\mathbf{x}} - \eta_{\underline{\mathbf{x}}}) \cdot (\underline{\mathbf{x}} - \eta_{\underline{\mathbf{x}}})^T \right] = E\left(\begin{bmatrix} \mathbf{x}_1 - \eta_1 \\ \mathbf{x}_2 - \eta_2 \\ \vdots \\ \mathbf{x}_n - \eta_n \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}_1 - \eta_1 & \mathbf{x}_2 - \eta_2 & \cdots & \mathbf{x}_n - \eta_n \end{bmatrix} \right) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

(گاهی آن را با $C_{\underline{\mathbf{x}}\underline{\mathbf{x}}}$ نیز نشان داده و آن را Auto Covariance هم می‌نامند.)

خواص ماتریس کواریانس:

۱. متقارن است، یعنی: $C = C^T$، زیرا: $\sigma_{ij} = \sigma_{ji}$.

۲. معیّن نامنفی (Nonnegative Definite) است، یعنی: برای هر بردار دلخواه $\underline{a}$ داریم:

$$\underline{a}^T C \underline{a} \geq 0$$

(اثبات این قضیه در تمرین‌ها خواسته شده است.)

از جمله خواص ماتریس‌های معین نامنفی این است که دترمینان آنها نامنفی است، پس داریم:

$$\det(C) \geq 0$$

ماتریس همبستگی $\underline{\mathbf{x}}$ (Correlation Matrix) طبق تعریف برابر است با:

$$R_{\underline{\mathbf{x}}} = E(\underline{\mathbf{x}} \cdot \underline{\mathbf{x}}^T)$$

از تعریف ماتریس‌های کواریانس و همبستگی نتیجه می‌شود:

$$C_{\underline{\mathbf{x}}} = R_{\underline{\mathbf{x}}} - \eta_{\underline{\mathbf{x}}} \cdot \eta_{\underline{\mathbf{x}}}^{T}$$

(شبیه به رابطهٔ $\sigma^2 = E(\mathbf{x}^2) - \left(E(\mathbf{x})\right)^2$)

در حالت کلی‌تر، می‌توانیم ماتریس‌های زیر را تعریف کنیم:

$$C_{\underline{\mathbf{x}}\underline{\mathbf{y}}} = \operatorname{cov}(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = E\left[(\underline{\mathbf{x}} - \eta_{\underline{\mathbf{x}}})(\underline{\mathbf{y}} - \eta_{\underline{\mathbf{y}}})^{T}\right] \quad : \text{Cross Covariance}$$

$$R_{\underline{\mathbf{x}}\underline{\mathbf{y}}} = E(\underline{\mathbf{x}} \cdot \underline{\mathbf{y}}^{T}) \quad : \text{Cross Correlation}$$

اگر $\mathbf{z} = \sum_{i=1}^{n} a_i \mathbf{x}_i$ باشد، خواهیم داشت:

$$E(\mathbf{z}) = \sum_{i=1}^{n} a_i E(\mathbf{x}_i) = \sum_{i=1}^{n} a_i \, \eta_i$$

$$\operatorname{var}(\mathbf{z}) = E\left[(\mathbf{z} - \eta_{\mathbf{z}})^2\right] = E\left[\left(\sum_{i=1}^{n} a_i (\mathbf{x}_i - \eta_i)\right)^2\right] = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (\mathbf{x}_i - \eta_i)(\mathbf{x}_j - \eta_j)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij}$$

یعنی اگر $\mathbf{z} = \underline{a}^T \underline{\mathbf{x}}$ باشد، داریم:

$$\eta_{\mathbf{z}} = \underline{a}^T \eta_{\underline{\mathbf{x}}}$$
$$\sigma_{\mathbf{z}}^2 = \underline{a}^T C_{\underline{\mathbf{x}}} \underline{a}$$

زیرا:

$$\sigma_{\mathbf{z}}^2 = E\left[(\mathbf{z}-\eta_{\mathbf{z}})^2\right] = E\left[\underline{a}^T(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\cdot(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^T \underline{a}\right] = \underline{a}^T E\left[(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\cdot(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^T\right]\underline{a} = \underline{a}^T C_{\underline{\mathbf{x}}} \underline{a}$$

$\mathbf{x}_i$ها را ناهمبسته گوییم، اگر برای هر $i \neq j$ داشته باشیم: $\sigma_{ij} = 0$؛ یعنی $C_{\underline{\mathbf{x}}}$ قطری باشد.

در این صورت خواهیم داشت:

$$\sigma_{\mathbf{z}}^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$ (اصل سوپر پوزیشن قدرت)

$\mathbf{x}_i$ها را متعامد گوییم، اگر برای هر $i \neq j$ داشته باشیم: $E(\mathbf{x}_i \mathbf{x}_j) = 0$؛ یعنی $R_{\underline{\mathbf{x}}}$ قطری باشد.

مشابه حالت دوبُعدی به سادگی می‌توان نشان داد که:

اگر $\mathbf{x}_i$ها مستقل باشند، ناهمبسته هم خواهند بود. همچنین $g_i(\mathbf{x}_i)$ها نیز ناهمبسته خواهند بود.

تابع مولد گشتاور و تابع مشخصه:

$$\Phi_{\underline{\mathbf{x}}}(s_1, s_2, \ldots, s_n) = E(e^{s_1\mathbf{x}_1 + s_2\mathbf{x}_2 + \cdots + s_n\mathbf{x}_n}) = E(e^{\sum_{i=1}^{n} s_i\mathbf{x}_i})$$

$$\Phi_{\underline{\mathbf{x}}}(j\omega_1, j\omega_2, \ldots, j\omega_n) = E(e^{j\omega_1\mathbf{x}_1 + j\omega_2\mathbf{x}_2 + \cdots + j\omega_n\mathbf{x}_n}) = E(e^{j\sum_{i=1}^{n} \omega_i\mathbf{x}_i})$$

به این ترتیب خواهیم داشت:

$$\frac{1}{j^{k_1+k_2+\cdots+k_n}} \frac{\partial^{k_1}\partial^{k_2}\cdots\partial^{k_n}}{\partial\omega_1\partial\omega_2\cdots\partial\omega_n} \Phi_{\underline{\mathbf{x}}}(j\omega_1, j\omega_2, \ldots, j\omega_n)\Big|_{\omega_1=\omega_2=\cdots=\omega_n=0} = E(\mathbf{x}_1^{k_1}\mathbf{x}_2^{k_2}\cdots\mathbf{x}_n^{k_n})$$

اگر $\mathbf{z} = \sum_{i=1}^{n} \mathbf{x}_i$ باشد، داریم:

$$\Phi_{\mathbf{z}}(j\omega) = E(e^{j\omega\mathbf{z}}) = E(e^{j\sum_{i=1}^{n}\omega\mathbf{x}_i}) = \Phi_{\underline{\mathbf{x}}}(j\omega, j\omega, \ldots, j\omega)$$

همچنین $\mathbf{x}_i$ها مستقلند، اگر و فقط اگر داشته باشیم:

$$\Phi_{\underline{\mathbf{x}}}(j\omega_1, j\omega_2, \ldots, j\omega_n) = \Phi_{\mathbf{x}_1}(j\omega_1)\Phi_{\mathbf{x}_2}(j\omega_2)\cdots\Phi_{\mathbf{x}_n}(j\omega_n) = \prod_{i=1}^{n}\Phi_{\mathbf{x}_i}(j\omega_i)$$

توابعی از $\mathbf{n}$ متغیر تصادفی:

اگر یک تابع از n متغیر تصادفی داشته باشیم، از چهار روشی که قبلاً در حالت دو متغیر تصادفی گفته بودیم، در اینجا نیز می‌توان برای به دست آوردن توزیع آن استفاده کرد.

مثال: متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$ و $\mathbf{x}_3$ مستقل بوده و همگی $N(0,1)$ هستند. اگر $\mathbf{v}=\mathbf{x}_1^2+\mathbf{x}_2^2+\mathbf{x}_3^2$ باشد، داریم:

$$F_{\mathbf{v}}(v)=P\{\mathbf{v}\le v\}=P\{\mathbf{x}_1^2+\mathbf{x}_2^2+\mathbf{x}_3^2\le v\}$$

برای $v<0$:

$$F_{\mathbf{v}}(v)=0$$

برای $v>0$:

$$F_{\mathbf{v}}(v)=\underset{\text{ناحیهٔ } x_1^2+x_2^2+x_3^2\le v}{\iiint} f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(x_1,x_2,x_3)\,dx_1dx_2dx_3$$

چون $\mathbf{x}_1$، $\mathbf{x}_2$ و $\mathbf{x}_3$ مستقل بوده و همگی $N(0,1)$ هستند، پس:

$$f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(x_1,x_2,x_3)=\frac{1}{(2\pi)^{\frac{3}{2}}}e^{-\frac{x_1^2+x_2^2+x_3^2}{2}}$$

(چنین توزیع‌هایی را که فقط تابع r (فاصله از مبدأ) هستند، متقارن کروی گویند. استقلال و تقارن کروی، نرمال بودن را نتیجه می‌دهند.)

پس:

$$F_{\mathbf{v}}(v)=\frac{1}{(2\pi)^{\frac{3}{2}}}\underset{\substack{\text{داخل کره‌ای} \\ \text{به شعاع } \sqrt{v}}}{\iiint} e^{-\frac{x_1^2+x_2^2+x_3^2}{2}}\,dx_1dx_2dx_3$$

با تبدیل به مختصات کروی خواهیم داشت:

$$\begin{cases} x_1 = r\sin\varphi\cos\theta \\ x_2 = r\sin\varphi\sin\theta \\ x_3 = r\cos\varphi \end{cases}$$

پس در نهایت داریم:

$$F_{\mathbf{v}}(v)=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_0^{2\pi}\int_0^{\pi}\int_0^{\sqrt{v}} e^{-\frac{r^2}{2}}\, r^2 \sin\varphi\, dr\, d\varphi\, d\theta=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_0^{2\pi} d\theta \int_0^{\pi}\sin\varphi\, d\varphi \int_0^{\sqrt{v}} r^2 e^{-\frac{r^2}{2}}\, dr$$

$$=\frac{(2\pi)(2)}{(2\pi)^{\frac{3}{2}}}\int_0^{\sqrt{v}} r^2 e^{-\frac{r^2}{2}}\, dr$$

$$f_{\mathbf{v}}(v)=\frac{d}{dv}F_{\mathbf{v}}(v)=\frac{2}{\sqrt{2\pi}}\cdot\frac{1}{2\sqrt{v}}\, v\, e^{-\frac{v}{2}}=\frac{1}{\sqrt{2\pi}}\, v^{\frac{1}{2}} e^{-\frac{v}{2}} : v>0$$

که توزیع χ^2 با سه درجه آزادی است.

اگر n تابع از n متغیر تصادفی داشته باشیم، یعنی:

$$\begin{cases} \mathbf{y}_1 = g_1(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \\ \mathbf{y}_2 = g_2(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \\ \vdots \\ \mathbf{y}_n = g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \end{cases} \rightarrow \quad \underline{\mathbf{y}} = \underline{g}(\underline{\mathbf{x}})$$

ابتدا دستگاه $\underline{y} = \underline{g}(\underline{x})$ را حل می‌کنیم. اگر ریشهٔ آن برابر با $\underline{h}(\underline{y})$ $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} h_1(\underline{y}) \\ h_2(\underline{y}) \\ \vdots \\ h_n(\underline{y}) \end{bmatrix} = \underline{h}(\underline{y})$ باشد، خواهیم داشت:

$$f_{\underline{\mathbf{y}}}(\underline{y}) = \frac{f_{\underline{\mathbf{x}}}(x_1, x_2, \ldots, x_n)}{\left|J(x_1, x_2, \ldots, x_n)\right|} = \frac{f_{\underline{\mathbf{x}}}\left(\underline{h}(\underline{y})\right)}{\left|J\left(\underline{h}(\underline{y})\right)\right|}$$

که:

$$J(x_1, x_2, \ldots, x_n) = \det\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}$$

یا معادلاً:

$$f_{\underline{y}}(\underline{y}) = f_{\underline{x}}\left(\underline{h}(\underline{y})\right) \left| \det \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_n}{\partial y_n} \end{bmatrix} \right|$$

(اگر چند ریشه وجود داشته باشد، عبارت فوق را در تمام ریشه‌ها حساب کرده و جمع می‌کنیم.)

مثال: اگر $\underline{\mathbf{y}} = A\underline{\mathbf{x}}$ باشد، یعنی: $\mathbf{y}_i = \sum_{j=1}^{n} a_{ij}\mathbf{x}_j : i = 1,2,\ldots,n$؛ $f_{\underline{\mathbf{y}}}$، $\eta_{\underline{\mathbf{y}}}$ و $C_{\underline{\mathbf{y}}}$ را پیدا کنید.

$$\underline{y}=A\underline{x} \quad \Rightarrow \quad \underline{x}=A^{-1}\underline{y}$$

$$J=\det(A)$$

$$f_{\underline{\mathbf{y}}}(\underline{y})=\frac{f_{\mathbf{x}}(A^{-1}\underline{y})}{|\det(A)|}$$

$$\eta_{\underline{\mathbf{y}}}=E(\underline{\mathbf{y}})=E(A\underline{\mathbf{x}})=A\eta_{\underline{\mathbf{x}}}$$

$$C_{\underline{\mathbf{y}}}=E\left[(\underline{\mathbf{y}}-\eta_{\underline{\mathbf{y}}})\cdot(\underline{\mathbf{y}}-\eta_{\underline{\mathbf{y}}})^T\right]=E\left[(A\underline{\mathbf{x}}-A\eta_{\underline{\mathbf{x}}})\cdot(A\underline{\mathbf{x}}-A\eta_{\underline{\mathbf{x}}})^T\right]=E\left[A(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\cdot(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^T A^T\right]=AC_{\underline{\mathbf{x}}}A^T$$

برای مثال داریم:

$$\begin{cases} \mathbf{y}_1 = \mathbf{x}_1 \\ \mathbf{y}_2 = \mathbf{x}_1 + \mathbf{x}_2 \\ \quad \vdots \\ \mathbf{y}_n = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n \end{cases} \Rightarrow \quad A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \Rightarrow \quad \det(A) = 1$$

$$\underline{y} = A\underline{x} \quad \Rightarrow \quad \begin{cases} x_1 = y_1 \\ x_2 = y_2 - y_1 \\ x_3 = y_3 - y_2 \\ \quad \vdots \\ x_n = y_n - y_{n-1} \end{cases}$$

$$f_{\underline{\mathbf{y}}}(\underline{y}) = f_{\underline{\mathbf{x}}}(y_1, y_2 - y_1, y_3 - y_2, \ldots, y_n - y_{n-1})$$

حال اگر $\mathbf{x}_i$ها مستقل باشند، $f_{\underline{\mathbf{x}}}(\underline{x}) = f_{\mathbf{x}_1}(x_1) f_{\mathbf{x}_2}(x_2) \cdots f_{\mathbf{x}_n}(x_n)$ و لذا:

$$f_{\underline{\mathbf{y}}}(\underline{y}) = f_{\mathbf{x}_1}(y_1) f_{\mathbf{x}_2}(y_2 - y_1) f_{\mathbf{x}_3}(y_3 - y_2) \cdots f_{\mathbf{x}_n}(y_n - y_{n-1})$$

متغیرهای تصادفی مشترکاً نرمال:

تعریف: متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ را مشترکاً نرمال گویند، هرگاه $\mathbf{z}=\sum_{i=1}^{n} a_i \mathbf{x}_i$ برای هر a_i نرمال باشد.

توابع $f_{\underline{\mathbf{x}}}$ و $\Phi_{\underline{\mathbf{x}}}$ برای متغیرهای تصادفی مشترکاً نرمال به صورت تابعی نمایی از یک فرم درجهٔ دوم هستند:

$$f_{\underline{\mathbf{x}}}(\underline{x})=\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(C_{\underline{\mathbf{x}}})}}\exp\left[-\frac{1}{2}(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^T C_{\underline{\mathbf{x}}}^{-1}(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\right]$$

یعنی $f_{\underline{\mathbf{x}}}$ به صورت $\alpha\, e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{ij}(x_i-\eta_i)(x_j-\eta_j)}$ است که γ_{ij}ها همان درایه‌های $C_{\underline{\mathbf{x}}}^{-1}$ هستند.

$$\Phi_{\underline{\mathbf{x}}}(j\underline{\omega})=e^{j\underline{\omega}^T\eta_{\underline{\mathbf{x}}}}\, e^{-\frac{1}{2}\underline{\omega}^T C_{\underline{\mathbf{x}}}\,\underline{\omega}}$$

یعنی $\Phi_{\underline{\mathbf{x}}}$ به صورت $e^{j\sum_{i=1}^{n}\omega_i\eta_i}\; e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\,\omega_i\,\omega_j}$ می‌باشد.

(این فرمول‌ها در حالت دوبُعدی به همان فرمول‌هایی که در فصل ۵ داشتیم تبدیل می‌شوند.)

ملاحظه می‌شود که در حالت مشترکاً نرمال، بردار میانگین و ماتریس کواریانس کلیهٔ خواص آماری $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ را مشخص می‌کنند.

قضیه: اگر هر یک از متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ نرمال بوده و مستقل نیز باشند، آنگاه مشترکاً نرمال خواهند بود.

(زیرا $\prod_{i=1}^{n} f_{\mathbf{x}_i}$ همان شکل فوق را خواهد داشت.)

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مشترکاً نرمال بوده و ناهمبسته نیز باشند، آنگاه مستقل خواهند بود.

اثبات:

$$\sigma_{ij} = 0 : i \neq j \quad \Rightarrow \quad C_{\underline{\mathbf{x}}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{bmatrix}$$

$$\Rightarrow f_{\underline{\mathbf{x}}}(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\prod_{i=1}^{n} \sigma_i^2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \eta_i)^2}{\sigma_i^2} \right\} = \prod_{i=1}^{n} f_{\mathbf{x}_i}(x_i)$$

(یا می‌توان نشان داد که: $\Phi_{\underline{\mathbf{x}}}(j\underline{\omega}) = \prod_{i=1}^{n} \Phi_{\mathbf{x}_i}(j\omega_i)$)

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مشترکاً نرمال بوده و $\underline{\mathbf{y}} = A\underline{\mathbf{x}}$ باشد، یعنی: $\mathbf{y}_i = \sum_{j=1}^{n} a_{ij}\,\mathbf{x}_j : i = 1,2,\ldots,m$، $\mathbf{y}_i$ها هم مشترکاً نرمال خواهند بود.

اثبات:

$$\mathbf{z} = \sum_{i=1}^{m} c_i \mathbf{y}_i = \sum_{i=1}^{m} c_i \sum_{j=1}^{n} a_{ij}\,\mathbf{x}_j = \sum_{j=1}^{n} \underbrace{\left(\sum_{i=1}^{m} c_i\, a_{ij} \right)}_{b_j} \mathbf{x}_j = \sum_{j=1}^{n} b_j\,\mathbf{x}_j \rightarrow$$ بنا به فرض دارای توزیع نرمال است

یا می‌توانیم به صورت برداری نیز نشان دهیم:

$$\mathbf{z} = C^T \underline{\mathbf{y}} = C^T (A\underline{\mathbf{x}}) = (C^T A)\underline{\mathbf{x}} = b^T \underline{\mathbf{x}} \rightarrow$$ بنا به فرض دارای توزیع نرمال است

کاربرد:

نویز گوسی ⟵ $\boxed{\text{مدار خطی}}$ ⟵ نویز گوسی

توزیع‌های شرطی:

مشابه حالت دوبُعدی در اینجا نیز داریم:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_k}(x_1,x_2,\ldots,x_k\,|\,\mathbf{x}_{k+1}=x_{k+1},\ldots,\mathbf{x}_n=x_n)=\frac{f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)}{f_{\mathbf{x}_{k+1}\cdots\mathbf{x}_n}(x_{k+1},\ldots,x_n)} \quad (*)$$

مثلاً داریم:

$$f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2\,|\,x_3,x_4)=\frac{f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3\mathbf{x}_4}(x_1,x_2,x_3,x_4)}{f_{\mathbf{x}_3\mathbf{x}_4}(x_3,x_4)}$$

با استفاده از رابطهٔ (*) می‌توان نشان داد که:

قاعدهٔ زنجیری:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)=f_{\mathbf{x}_1}(x_1)f_{\mathbf{x}_2}(x_2\,|\,x_1)f_{\mathbf{x}_3}(x_3\,|\,x_2,x_1)\cdots f_{\mathbf{x}_n}(x_n\,|\,x_1,x_2,\ldots,x_{n-1})$$

حذف متغیر سمت چپ در چگالی شرطی:

می‌دانیم که اگر $f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2)$ را داشته باشیم و بخواهیم $f_{\mathbf{x}_1}$ را به دست آوریم، کافی است روی $f_{\mathbf{x}_1\mathbf{x}_2}$ از $-\infty$ تا $+\infty$ نسبت به x_2 انتگرال بگیریم. در چگالی شرطی نیز به همین ترتیب عمل می‌کنیم. یعنی مثلاً اگر بخواهیم از $f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3)$ مقدار $f_{\mathbf{x}_1}(x_1|x_3)$ را به دست آوریم (توجه کنید که عین خواص pdf معمولی برقرار است)، خواهیم داشت:

$$f_{\mathbf{x}_1}(x_1|x_3)=\int_{-\infty}^{+\infty} f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3)\,dx_2$$

(برای اثبات از تعریف چگالی شرطی و رابطهٔ چگالی مشترک حاشیه‌ای استفاده کنید.)

حذف متغیر سمت راست در چگالی شرطی:

اگر بخواهیم از $f_{\mathbf{x}_1}(x_1|x_2,x_3)$ به $f_{\mathbf{x}_1}(x_1|x_3)$ برسیم، داریم:

$$f_{\mathbf{x}_1}(x_1|x_3)=\int_{-\infty}^{+\infty} f_{\mathbf{x}_1}(x_1|x_2,x_3)f_{\mathbf{x}_2}(x_2|x_3)dx_2$$

زیرا:

$$f_{\mathbf{x}_1}(x_1|x_2,x_3)f_{\mathbf{x}_2}(x_2|x_3)=f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3)$$

لذا:

$$f_{\mathbf{x}_1}(x_1|x_3)=\int_{-\infty}^{+\infty} f_{\mathbf{x}_1}(x_1|x_2,x_3)f_{\mathbf{x}_2}(x_2|x_3)dx_2$$ (معادلهٔ چاپمن- کولموگروف)

اگر $\mathbf{x}_i$ها مستقل باشند، چگالی مشروط بعضی از آنها بر بعضی دیگر با چگالی غیرمشروط یکسان است. مثلاً:

$$f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3,x_4)=f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2)$$

متوسط مشروط:

$$E(\mathbf{y}|\mathbf{x}_1=x_1,\mathbf{x}_2=x_2,\dots,\mathbf{x}_n=x_n)=E(\mathbf{y}|x_1,x_2,\dots,x_n)=\int_{-\infty}^{+\infty} y\, f_{\mathbf{y}}(y|x_1,x_2,\dots,x_n)\,dy$$

(که تابعی از x_1، x_2، ... و x_n است، نه y.)

$$E(g(\mathbf{y})|x_1,x_2,\dots,x_n)=\int_{-\infty}^{+\infty} g(y)\, f_{\mathbf{y}}(y|x_1,x_2,\dots,x_n)\,dy$$

(که تابعی از x_1، x_2، ... و x_n است، نه y.)

حال اگر فرض کنیم:

$$E(\mathbf{y}|\underline{x})=\Phi(\underline{x})$$

متناظر با این تابع می‌توانیم متغیر تصادفی $\Phi(\underline{\mathbf{x}})$ را تعریف کنیم:

$$E(\mathbf{y}|\underline{\mathbf{x}})=\Phi(\underline{\mathbf{x}})$$

مشابه قبل به سادگی می‌توان نشان داد که:

$$\mathrm{E}\left[\mathrm{E}(\mathbf{y}\,|\,\underline{\mathbf{x}})\right]=\mathrm{E}(\mathbf{y})$$

و اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، خواهیم داشت:

$$\mathrm{E}(\mathbf{y}\,|\,\underline{\mathbf{x}})=\mathrm{E}(\mathbf{y})$$

$\mathrm{E}(\mathbf{y}\,|\,\underline{\mathbf{x}})$ در واقع تخمین بهینه (به مفهوم mse) برای $\mathbf{y}$ بر مبنای مشاهدهٔ $\underline{\mathbf{x}}$ است.

برخی کاربردهای توزیع شرطی و دنبالهٔ متغیرهای تصادفی:

۱. مجموع‌های تصادفی (Random Sums):

فرض کنید $\mathbf{y}=\sum_{i=1}^{N}\mathbf{x}_i$ باشد که در آن، $\mathbf{x}_i$ها متغیرهای تصادفی مستقل بوده و نیز داریم:

$$\forall i: E(\mathbf{x}_i)=\eta_{\mathbf{x}}\ ,\ \mathrm{var}(\mathbf{x}_i)=\sigma_{\mathbf{x}}^2$$

از طرف دیگر، خود $\mathbf{N}$ نیز یک متغیر تصادفی با میانگین $\eta_{\mathbf{N}}$ و واریانس $\sigma_{\mathbf{N}}^2$ باشد و $\mathbf{x}_i$ها از $\mathbf{N}$ نیز مستقل باشند.

مثلاً $\mathbf{y}$ می‌تواند سود روزانهٔ یک مغازه، $\mathbf{x}_i$ سود حاصل از هر مشتری و $\mathbf{N}$ تعداد مشتری‌هایی که در روز مراجعه می‌کنند، باشد.
یا در مسألهٔ Multipath Fading یا انرژی کل الکترون‌های منتشر شده از یک مادهٔ رادیواکتیو ($\mathbf{N}$ دارای توزیع پواسون).

حال می‌خواهیم میانگین و واریانس $\mathbf{y}$ را به دست آوریم:

$$E(\mathbf{y}) = E\left[E(\mathbf{y}|\mathbf{N})\right]$$

$$E(\mathbf{y}|\mathbf{N}=n) = E\left(\sum_{i=1}^{n}(\mathbf{x}_i|n)\right) = \sum_{i=1}^{n}E(\mathbf{x}_i|n) = \sum_{i=1}^{n}E(\mathbf{x}_i)$$

ولی داشتیم:

$$\forall i : E(\mathbf{x}_i) = \eta_{\mathbf{x}}$$

پس داریم:

$$E(\mathbf{y}|n) = n\eta_{\mathbf{x}}$$

$$\Rightarrow E(\mathbf{y}|\mathbf{N}) = \mathbf{N}\eta_{\mathbf{x}}$$

$$\Rightarrow E(\mathbf{y}) = E\left[E(\mathbf{y}|\mathbf{N})\right] = E(\mathbf{N}\eta_{\mathbf{x}}) = \eta_{\mathbf{x}}E(\mathbf{N}) = \eta_{\mathbf{x}}\eta_{\mathbf{N}}$$

$$E(\mathbf{y}^2 \mid n) = E\left[\left(\sum_{i=1}^{n} \mathbf{x}_i\right)^2\right] = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{x}_i \mathbf{x}_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(\mathbf{x}_i \mathbf{x}_j)$$

به دلیل استقلال $\mathbf{x}_i$ها، برای $i \neq j$ داریم:

$$E(\mathbf{x}_i \mathbf{x}_j) = E(\mathbf{x}_i) E(\mathbf{x}_j)$$

در نتیجه:

$$E(\mathbf{y}^2 \mid n) = \sum_{i=1}^{n} E(\mathbf{x}_i^2) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} E(\mathbf{x}_i) E(\mathbf{x}_j)$$

ولی داشتیم:

$$\forall i : \mathrm{var}(\mathbf{x}_i) = \sigma_{\mathbf{x}}^2$$

پس خواهیم داشت:

$$E(\mathbf{y}^2 \mid n) = n(\sigma_{\mathbf{x}}^2 + \eta_{\mathbf{x}}^2) + (n^2 - n)\eta_{\mathbf{x}}^2 = n\sigma_{\mathbf{x}}^2 + n^2 \eta_{\mathbf{x}}^2$$

$$\Rightarrow E(\mathbf{y}^2 \mid \mathbf{N}) = \mathbf{N}\sigma_{\mathbf{x}}^2 + \mathbf{N}^2 \eta_{\mathbf{x}}^2$$

$$E(\mathbf{y}^2) = E\left[E(\mathbf{y}^2 \mid \mathbf{N})\right] = \sigma_{\mathbf{x}}^2 E(\mathbf{N}) + \eta_{\mathbf{x}}^2 E(\mathbf{N}^2) = \sigma_{\mathbf{x}}^2 \eta_{\mathbf{N}} + \eta_{\mathbf{x}}^2 (\sigma_{\mathbf{N}}^2 + \eta_{\mathbf{N}}^2)$$

$$\Rightarrow \sigma_{\mathbf{y}}^2 = E(\mathbf{y}^2) - \left(E(\mathbf{y})\right)^2 = \sigma_{\mathbf{x}}^2 \eta_{\mathbf{N}} + \eta_{\mathbf{x}}^2 \sigma_{\mathbf{N}}^2 + \eta_{\mathbf{x}}^2 \eta_{\mathbf{N}}^2 - \eta_{\mathbf{x}}^2 \eta_{\mathbf{N}}^2 = \sigma_{\mathbf{x}}^2 \eta_{\mathbf{N}} + \eta_{\mathbf{x}}^2 \sigma_{\mathbf{N}}^2$$

۲. نمونه‌برداری:

متغیر تصادفی $\mathbf{x}$ را طول قد مردان ایرانی در نظر گرفته و فرض کنیم (بنابر استدلالات یا مشاهدهٔ طولانی) که دارای چگالی $f_{\mathbf{x}}$ و توزیع $F_{\mathbf{x}}$ باشد (مثلاً $N(170,10)$). حال مردی را به عنوان نمونه (به طور کاملاً تصادفی) انتخاب کنیم و از شما سؤال کنیم احتمال اینکه قد این مرد کوچکتر از x باشد، چقدر است؟

خواهید گفت:

$$P\{\mathbf{x}_1 \le x\} = F_{\mathbf{x}}(x)$$

به همین ترتیب برای سایر نمونه‌ها نیز داریم:

$$F_{\mathbf{x}_i}(x) = F_{\mathbf{x}}(x)\ ,\ f_{\mathbf{x}_i}(x) = f_{\mathbf{x}}(x)$$

از سوی دیگر، با فرض استقلال آزمایشها، $\mathbf{x}_i$ها مستقلند (تعداد نمونه‌ها باید نسبت به تعداد افراد جامعه کم باشد یا نمونه‌برداری با جایگزینی باشد).

یا مثلاً اگر متغیر تصادفی $\mathbf{x}$، قطر پیچ‌های تولیدی یک کارخانه باشد و فرض کنیم دارای چگالی $f_{\mathbf{x}}$ باشد، این مدلی است که طبق فرض برای کلیهٔ $\mathbf{x}$ها صادق است. پس اگر $\mathbf{x}_i$ قطر پیچ نمونهٔ i ام باشد، داریم:

$$f_{\mathbf{x}_i}(x) = f_{\mathbf{x}}(x)$$

و $\mathbf{x}_i$ها (با فرض استقلال آزمایشها) مستقلند.

به طور کلی، متغیرهای تصادفی که مستقل و دارای توزیع یکسان باشند، i.i.d. (Independent Identically Distributed) نامیده می‌شوند.

اگر $\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$ و $\mathbf{x}_i$ها i.i.d. باشند، داریم:

$$f_{\underline{\mathbf{x}}}(\underline{x}) \underset{\boxed{\text{استقلال}}}{=} f_{\mathbf{x}_1}(x_1)\cdots f_{\mathbf{x}_n}(x_n) \underset{\boxed{\text{یکسان بودن توزیع}}}{=} f_{\mathbf{x}}(x_1)\cdots f_{\mathbf{x}}(x_n)$$

با نمونه‌برداری مکرر از یک متغیر تصادفی در یک جامعه (Population) (مثلاً جامعهٔ پیچ‌ها در مثال بالا)، n متغیر تصادفی i.i.d. به دست می‌آیند. n را اندازهٔ نمونه (Sample Size) می‌گویند.

چون $\mathbf{x}_i$ها مستقلند، پس:

$$f_{\mathbf{x}_i}(x) = f_{\mathbf{x}}(x)$$

میانگین جامعه : $\Rightarrow E(\mathbf{x}_i) = E(\mathbf{x}) = \eta$

واریانس جامعه : $\Rightarrow \mathrm{var}(\mathbf{x}_i) = \mathrm{var}(\mathbf{x}) = \sigma^2$

بحث نمونه‌برداری نقش اساسی در آمار دارد.

میانگین نمونه (Sample Mean):

طبق تعریف، برابر است با:

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$$

داریم:

$$E(\bar{\mathbf{x}}) = \frac{1}{n}\sum_{i=1}^{n} E(\mathbf{x}_i) = \frac{1}{n}\cdot n\eta = \eta$$

$$\sigma_{\bar{\mathbf{x}}}^2 = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_{\mathbf{x}_i}^2 = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

پس هر چقدر n زیادتر شود، مقدار $\bar{\mathbf{x}}$ به η واقعی نزدیکتر خواهد بود. $\bar{\mathbf{x}}$ میانگین نمونه است، در حالی که η میانگین جامعه می‌باشد (پارامتر مدل مفروض).

با توجه به قضیهٔ چبیشف داریم:

$$P\{|\bar{\mathbf{x}}-\eta|<\varepsilon\}\geq 1-\frac{\sigma_{\bar{\mathbf{x}}}^2}{\varepsilon^2}$$

یعنی:

$$P\{|\bar{\mathbf{x}}-\eta|<\varepsilon\}\geq 1-\frac{\sigma^2}{n\varepsilon^2}$$

پس این احتمال با افزایش n به یک نزدیک می‌شود.

به همین ترتیب می‌توان در مورد واریانس نمونه بحث کرد که بعداً آن را خواهیم دید.

حال فرض کنید که یک آزمایش تصادفی را n بار انجام دهیم ($\mathbf{x}_i$ها طبق تعریفی که داشتیم تعریف شده باشند) و مقادیر x_i در این آزمایشها ظاهر شوند $(i=1,2,\dots,n)$. این اعداد را به ترتیب صعودی مرتب می‌کنیم:

$$x_{r_1}\leq x_{r_2}\leq\cdots\leq x_{r_n}$$

اکنون نام مرتب شدهٔ آنها را y_i می‌نامیم، یعنی: $y_i=x_{r_i}$. پس خواهیم داشت:

$$y_1\leq y_2\leq\cdots\leq y_n$$

مثلاً اگر اعداد x_i به ترتیب زیر به دست آمده باشند، داریم:

$$\begin{cases} x_1 = 12 \\ x_2 = 7 \\ x_3 = 13 \\ x_4 = 8 \\ x_5 = 9 \end{cases} \Rightarrow \begin{cases} y_1 = x_{r_1} = x_2 = 7 \\ y_2 = x_{r_2} = x_4 = 8 \\ y_3 = x_{r_3} = x_5 = 9 \\ y_4 = x_{r_4} = x_1 = 12 \\ y_5 = x_{r_5} = x_3 = 13 \end{cases}$$

متغیرهای تصادفی $\mathbf{y}_i$ را آماره‌های رتبهٔ (آمارگان رتبهٔ) $\mathbf{x}_i$ها می‌نامند (آماره‌های رتبه کاربرد بسیاری در آشکارسازی و تخمین پارامتری دارند و در رادار و جنگ الکترونیک مورد استفاده قرار می‌گیرند).

<u>آماره (Statistic):</u>

طبق تعریف، تابعی از یک یا چند متغیر تصادفی را که به پارامتر نامعلومی بستگی نداشته باشند، آماره گویند. در بحث نمونه‌برداری، تابعی از صرفاً بردار متغیرهای تصادفی نمونهٔ $\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$ را آماره گویند. مثلاً آمارهٔ رتبه یک نوع آماره است (زیرا تابعی از $\underline{\mathbf{x}}$ می‌باشد).

به عنوان حالت خاص برای $\mathbf{y}_1$ و $\mathbf{y}_n$ داریم:

$$\mathbf{y}_1 = \mathbf{x}_{min} = \min(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$
$$\mathbf{y}_n = \mathbf{x}_{max} = \max(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$

پس:

$$\begin{aligned} F_{\mathbf{y}_1}(y) &= P\{\mathbf{y}_1 \le y\} = P\{\mathbf{x}_{min} \le y\} = 1 - P\{\mathbf{x}_{min} > y\} = 1 - P\{\mathbf{x}_1 > y, \mathbf{x}_2 > y, \ldots, \mathbf{x}_n > y\} \\ &= 1 - P\{\mathbf{x}_1 > y\}P\{\mathbf{x}_2 > y\}\cdots P\{\mathbf{x}_n > y\} \\ &= 1 - [1 - F_{\mathbf{x}}(y)]^n \end{aligned}$$
$$\Rightarrow f_{\mathbf{y}_1}(y) = n[1 - F_{\mathbf{x}}(y)]^{n-1} f_{\mathbf{x}}(y)$$

$$\begin{aligned} F_{\mathbf{y}_n}(y) &= P\{\mathbf{y}_n \le y\} = P\{\mathbf{x}_{max} \le y\} = P\{\mathbf{x}_1 \le y, \mathbf{x}_2 \le y, \ldots, \mathbf{x}_n \le y\} \\ &= P\{\mathbf{x}_1 \le y\}P\{\mathbf{x}_2 \le y\}\cdots P\{\mathbf{x}_n \le y\} \\ &= [F_{\mathbf{x}}(y)]^n \end{aligned}$$
$$\Rightarrow f_{\mathbf{y}_n}(y) = n[F_{\mathbf{x}}(y)]^{n-1} f_{\mathbf{x}}(y)$$

(به کمک pdf نیز می‌توان روابط بالا را به دست آورد.)

در حالت کلی داریم:

$$f_{\mathbf{y}_k}(y)\,dy = P\{y < \mathbf{y}_k \le y+dy\}$$

در صورتی این واقعه را خواهیم داشت که از n تا x_i، $k-1$تا کوچکتر از y، $n-k$تا بزرگتر از $y+dy$ و یکی بین y و $y+dy$ باشد.

یادآوری توزیع چندجمله‌ای: احتمال اینکه در n بار آزمایش، واقعهٔ A_1، k_1بار، واقعهٔ A_2، k_2بار و واقعهٔ A_3، k_3بار اتفاق افتد ($k_1+k_2+k_3=n$ و $P(A_i)=p_i$) برابر است با:

$$p = \frac{n!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}$$

در نتیجه داریم:

$$f_{\mathbf{y}_k}(y)\,dy = \frac{n!}{(k-1)!(n-k)!1!}\left[F_{\mathbf{x}}(y)\right]^{k-1}\left[1-F_{\mathbf{x}}(y+dy)\right]^{n-k} f_{\mathbf{x}}(y)\,dy$$

$$dy \to 0 \quad \Rightarrow \quad f_{\mathbf{y}_k}(y) = \frac{n!}{(k-1)!(n-k)!}\left[F_{\mathbf{x}}(y)\right]^{k-1}\left[1-F_{\mathbf{x}}(y)\right]^{n-k} f_{\mathbf{x}}(y)$$

برای $k=1$، توزیع $\mathbf{x}_{\min}$ و برای $k=n$، توزیع $\mathbf{x}_{\max}$ به دست می‌آید:

$$f_{\mathbf{x}_{\min}}(y)=n\left[1-F_{\mathbf{x}}(y)\right]^{n}f_{\mathbf{x}}(y) \quad , \quad 1-F_{\mathbf{x}_{\min}}(y)=\left[1-F_{\mathbf{x}}(y)\right]^{n}$$

$$f_{\mathbf{x}_{\max}}(y)=n\left[F_{\mathbf{x}}(y)\right]^{n-1}f_{\mathbf{x}}(y) \quad , \quad F_{\mathbf{x}_{\max}}(y)=\left[F_{\mathbf{x}}(y)\right]^{n}$$

مجموع متغیرهای تصادفی مستقل:

می‌دانیم که اگر $\mathbf{z} = \mathbf{x}_1 + \mathbf{x}_2$ بوده و $\mathbf{x}_1$ و $\mathbf{x}_2$ مستقل باشند، آنگاه داریم:

$$\Phi_{\mathbf{z}} = \Phi_{\mathbf{x}_1}\Phi_{\mathbf{x}_2} \quad , \quad f_{\mathbf{z}} = f_{\mathbf{x}_1} * f_{\mathbf{x}_2}$$

همچنین دیدیم که اگر $\mathbf{z} = \sum_{i=1}^{n} \mathbf{x}_i$ باشد، آنگاه:

$$\Phi_{\mathbf{z}}(j\omega) = \Phi_{\underline{\mathbf{x}}}(j\omega, \ldots, j\omega)$$

و اگر $\mathbf{x}_i$ها مستقل باشند، خواهیم داشت:

$$\Phi_{\mathbf{z}}(j\omega) = \Phi_{\mathbf{x}_1}(j\omega)\Phi_{\mathbf{x}_2}(j\omega)\cdots\Phi_{\mathbf{x}_n}(j\omega)$$

$$f_{\mathbf{z}} = f_{\mathbf{x}_1} * f_{\mathbf{x}_2} * \cdots * f_{\mathbf{x}_n}$$

(با توجه به اینکه عمل کانولوشن دارای خاصیت جابجایی و شرکت‌پذیری است.)

مثال: اگر $\mathbf{x}_i$ها مستقل بوده و هر یک دارای توزیع پواسن با پارامتر λ_i باشند، داریم:

$$\Phi_{\mathbf{z}}(j\omega)=\prod_{i=1}^{n}\Phi_{\mathbf{x}_i}(j\omega)=\prod_{i=1}^{n}e^{\lambda_i(e^{j\omega}-1)}=e^{\lambda(e^{j\omega}-1)}$$

که: $\lambda=\sum_{i=1}^{n}\lambda_i$ (یعنی مجموعشان هم دارای توزیع پواسن است، اما با پارامتر $\sum_{i=1}^{n}\lambda_i$).

اگر $\mathbf{x}_i$ها .i.i.d باشند، یعنی: $\forall x: f_{\mathbf{x}_1}(x) = f_{\mathbf{x}_2}(x) = \cdots = f_{\mathbf{x}_n}(x) = f_{\mathbf{x}}(x)$، در این صورت خواهیم داشت:

$$f_{\mathbf{z}} = f_{\mathbf{x}} * f_{\mathbf{x}} * \cdots * f_{\mathbf{x}}$$

$$\Phi_{\mathbf{z}}(\omega) = \Phi_{\mathbf{x}}^{n}(\omega)$$

$$E(\mathbf{z}) = n\,E(\mathbf{x})$$

$$\sigma_{\mathbf{z}}^2 = n\sigma_{\mathbf{x}}^2$$

مثال: اگر $\mathbf{x}_i$ها مستقل بوده و همگی دارای توزیع نمایی باشند، یعنی: $f_{\mathbf{x}_i}(x) = \lambda e^{-\lambda x} u(x)$ (یعنی .i.i.d باشند) و $\mathbf{z} = \sum_{i=1}^{n} \mathbf{x}_i$ باشد، آنگاه $f_{\mathbf{z}}$ چه خواهد بود؟ مثلاً اگر سیستمی از n المان به صورت Standby استفاده کند و خرابی این المان‌ها از هم مستقل باشد و برای همگی داشته باشیم: $f_{\mathbf{x}}(x) = \lambda e^{-\lambda x} u(x)$، تابع چگالی عمر سیستم چه خواهد بود؟

$$\Phi_{\mathbf{x}}(j\omega) = \frac{\lambda}{\lambda - j\omega} \quad \Rightarrow \quad \Phi_{\mathbf{z}}(j\omega) = \left(\frac{\lambda}{\lambda - j\omega}\right)^n$$

با استفاده از جداول تبدیل فوریه داریم:

$$\frac{x^{n-1}}{(n-1)!} e^{-\lambda x} u(x) \xleftrightarrow{\ \mathfrak{I}\ } \frac{1}{(\lambda + j\omega)^n}$$

لذا:

$$f_{\mathbf{z}}(z) = \lambda^n \frac{z^{n-1}}{(n-1)!} e^{-\lambda z} u(z)$$

که همان توزیع ارلانگ است (توزیع زمان لازم برای وقوع n واقعهٔ کاملاً تصادفی یا به عبارتی مجموع n نمایی).

برای توزیع گاما $E(\mathbf{z}) = \frac{r}{\lambda}$ است، در نتیجه داریم:

$$E(\mathbf{z}) = \frac{n}{\lambda}$$

پس همان طور که انتظار داشتیم: $E(\mathbf{z}) = n\,E(\mathbf{x})$؛ یعنی MTTF، n برابر شده است.

اگر این المان‌ها از نظر نقش در خرابی یا صحت سیستم به صورت موازی بودند، خواهیم داشت:

$$\mathbf{z} = \max(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

$$F_{\mathbf{z}}(z) = F_{\mathbf{x}}^{n}(z)$$

$$f_{\mathbf{z}}(z) = n\, F_{\mathbf{x}}^{n-1}(z)\, f_{\mathbf{x}}(z)$$

برای $f_{\mathbf{x}}(x) = \lambda e^{-\lambda x} u(x)$ می‌توانید نشان دهید که:

$$E(\mathbf{z}) = \frac{1}{\lambda}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)$$

مثلاً برای $n = 4$، مقدار MTTF، 2.1 برابر می‌شود.

در صورتی که این المان‌ها سری باشند، خواهیم داشت:

$$\mathbf{z}=\min(\mathbf{x}_1,\ldots,\mathbf{x}_n)$$

$$F_{\mathbf{z}}(z)=1-(1-F_{\mathbf{x}}(z))^n=1-(e^{-\lambda z})^n$$

$$f_{\mathbf{z}}(z)=n\lambda e^{-n\lambda z}u(z)$$

(توجه کنید که مقدار می‌نیمم تعدادی از متغیرهای تصادفی نمایی و .i.i.d، خود نیز نمایی است.)

پس:

$$E(\mathbf{z})=\frac{1}{n\lambda}$$

یعنی MTTF، $\frac{1}{n}$ برابر شده است.

مثال: اگر متغیر تصادفی $\mathbf{x}$ به صورت زیر تعریف شود:

$$\mathbf{x}(\omega)=\begin{cases}1 & \omega\in A\\ 0 & \text{otherwise}\end{cases} \quad , \quad P(A)=p$$

مثلاً A واقعهٔ آمدن شیر در پرتاب سکه باشد، $\mathbf{x}$ دارای توزیع برنولی (دو جمله‌ای با $n=1$) است. پس:

$$E(\mathbf{x})=p \quad , \quad \sigma_{\mathbf{x}}^2=pq$$

حال با n بار تکرار این آزمایش (نمونه‌برداری از $\mathbf{x}$) و با فرض مستقل بودن آزمایش‌ها، $\mathbf{x}_i$های i.i.d. حاصل می‌شوند.

پس اگر تعریف کنیم:

$$\mathbf{z}=\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_n$$

$\mathbf{z}$ تعداد شیرهای به دست آمده در n بار پرتاب سکه خواهد بود. یعنی:

$$\mathbf{z}\sim \text{Binomial}(n,p)$$

و خواهیم داشت:

$$E(\mathbf{z})=n\,E(\mathbf{x})=np$$

$$\sigma_{\mathbf{z}}^2=n\sigma_{\mathbf{x}}^2=npq$$

(می‌بینیم که این روش خیلی ساده‌تر از روش مستقیم و حتی خیلی ساده‌تر از روش تابع مشخصه است.)

مثال: متغیر تصادفی $\mathbf{x}$ را برابر تعداد شکست‌ها تا حصول اولین موفقیت تعریف می‌کنیم. می‌دانیم که $\mathbf{x}$ دارای توزیع هندسی است (توزیع دو جمله‌ای منفی با $r=1$). در تمرین سری ۵ نشان دادید (از راه تابع مشخصه) که:

$$E(\mathbf{x})=\frac{q}{p} \quad , \quad \sigma_{\mathbf{x}}^2=\frac{q}{p^2} \quad , \quad \Phi_{\mathbf{x}}(j\omega)=\frac{p}{1-q\,e^{j\omega}}$$

حال اگر تعریف کنیم:

$$\mathbf{z}=\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_r$$

که همهٔ $\mathbf{x}_i$ها دارای توزیع هندسی بوده و مستقلند، آنگاه $\mathbf{z}$ تعداد شکست‌ها تا حصول r امین موفقیت خواهد بود. یعنی:

$$\mathbf{z}\sim \text{Negative Binomial}(r,p)$$

پس:

$$E(\mathbf{z})=\frac{rq}{p} \quad , \quad \sigma_{\mathbf{z}}^2=\frac{rq}{p^2}$$

در حالی که محاسبهٔ این دو از راه مستقیم یا تابع مشخصه خیلی مشکلتر است.

همچنین در مورد $\Phi_{\mathbf{z}}(j\omega)$ داریم:

$$\Phi_{\mathbf{z}}(j\omega)=\left(\frac{p}{1-q\,e^{j\omega}}\right)^r$$

در حالی که به دست آوردن مستقیم تابع مشخصه مشکلتر است.

قضیهٔ حد مرکزی (Central Limit Theorem):

طبق قضیهٔ حد مرکزی، اگر $\mathbf{x}_i$ها مستقل بوده و $\mathbf{y}=\sum_{i=1}^{n}\mathbf{x}_i$ باشد، وقتی $n\to+\infty$، توزیع $\mathbf{y}$ به توزیع نرمال میل می‌کند (با شرایط سهلی)، ولو $\mathbf{x}_i$ها نرمال نباشند. این در واقع بیانگر خاصیتی از کانولوشن است که کانولوشن تعداد زیادی توابع مثبت به گوسی میل می‌کند. با توجه به CLT معلوم می‌شود که چرا بسیاری از پدیده‌ها در جهان خارج توزیع تقریباً نرمال دارند. اصولاً هرگاه پدیده‌ای تحت تأثیر عوامل متعدد تصادفی باشد (قد یک فرد، خطا در اندازه‌گیری، ولتاژ نویز حرارتی و ...) تقریباً نرمال خواهد بود.

قضیهٔ حد مرکزی لیاپانوف:

اگر $\mathbf{y}=\sum_{i=1}^{n}\mathbf{x}_i$ بوده و $\mathbf{x}_i$ها مستقل باشند و $E(|\mathbf{x}_i-\eta_i|^3)<+\infty$ (یا $E(|\mathbf{x}_i|^3)<+\infty$) و $\sum_{i=1}^{n}\sigma_i^2\xrightarrow[n\to+\infty]{}+\infty$، آنگاه برای $\mathbf{z}=\dfrac{\mathbf{y}-\eta_y}{\sigma_y}$ داریم:

$$F_{\mathbf{z}}(z)\xrightarrow[n\to+\infty]{}G(z)$$

در قضیهٔ لیاپانوف لزوماً توزیع $\mathbf{x}_i$ها یکسان نیست (لزوماً .i.i.d نیستند). صورت دیگری از قضیهٔ حد مرکزی داریم که صرفاً برای $\mathbf{x}_i$های .i.i.d است.

قضیهٔ حد مرکزی لیندبرگ- لوی:

اگر $\mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i$ بوده و $\mathbf{x}_i$ها .i.i.d با واریانس محدود $\sigma^2 < +\infty$ باشند، آنگاه برای $\mathbf{z} = \frac{\mathbf{y} - \eta_y}{\sigma_y}$ داریم:

$$F_{\mathbf{z}}(z) \xrightarrow[n \to +\infty]{} G(z)$$

قضیهٔ لیندبرگ- لوی حالت خاصی از قضیهٔ لیاپانوف نمی‌باشد. چون در اینجا دیگر شرط $E\left[(\mathbf{x}_i - \eta_i)^3\right] < +\infty$ لازم نیست و شرط ضعیفتر $\sigma^2 < +\infty$ جایگزین آن شده است (شرط دیگر قضیهٔ لیاپانوف، یعنی $\sum_{i=1}^{n} \sigma_i^2 \xrightarrow[n \to +\infty]{} +\infty$، مسلماً ارضاء می‌شود).

همچنین قضیهٔ لیندبرگ- فلر را داریم که برای حالت کلی (x_iهای نه لزوماً .i.i.d) شرط لازم و کافی گوسی بودن توزیع حدی را ارائه می‌کند که تعبیر با تسامحش این است که $\forall j: \lim_{n\to+\infty} \frac{\sigma_j^2}{\sum_{i=1}^{n}\sigma_i^2} = 0$. یعنی هیچ جمله‌ای غالب نباشد و در مورد همهٔ جمله‌ها، واریانس آن در مقابل واریانس کل ناچیز باشد. توجه کنید که برقراری شرایط قضیهٔ لیاپانوف، برقراری شرایط قضیهٔ لیندبرگ- فلر را نتیجه می‌دهد.

وقتی n بی‌نهایت نبوده ولی بزرگ باشد هم می‌توانیم نرمال را به عنوان تقریب به کار ببریم. اگر توزیع $\mathbf{x}$ هموار باشد، حتی برای $n=5$، بسیار به توزیع نرمال نزدیک خواهیم بود. در اغلب کاربردها $n=30$ کاملاً کافی است (در کتاب، $n=3$ را برای توزیع یکنواخت به کار برده و با نرمال مقایسه کرده است).

مثال: جعبه ای شامل ۱۰۰ مقاومت $1k\Omega \pm 5\%$ در اختیار داریم. اگر توزیع مقدار مقاومت‌ها را یکنواخت فرض کنیم، احتمال اینکه مجموع مقدار این مقاومت‌ها در محدودهٔ $100k\Omega \pm 0.5\%$ باشد، چیست؟

$$\mathbf{y} = \sum_{i=1}^{100} \mathbf{x}_i$$

$$\sigma_{\mathbf{x}}^2 = \frac{(b-a)^2}{12} = \frac{(100)^2}{12} < +\infty$$

پس شرط قضیهٔ لیندبرگ- لوی برقرار است (اگر توزیع را نمی‌دانستیم و فقط همین واریانس را می‌دانستیم هم کافی بود).

$$E(\mathbf{x}) = 1000 \quad \Rightarrow \quad E(\mathbf{y}) = 100000$$

$$\sigma_{\mathbf{y}}^2 = n\sigma_{\mathbf{x}}^2 = 100\frac{(100)^2}{12} = \frac{10^6}{12}$$

$$P\{99500 < \mathbf{y} < 100500\} = G\left(\frac{100500-100000}{\sqrt{\frac{10^6}{12}}}\right) - G\left(\frac{99500-100000}{\sqrt{\frac{10^6}{12}}}\right) = 2G(1.732) - 1 = 0.917$$

در بیان قضیه‌های فوق از تابع توزیع انباشته استفاده شد تا برای حالت گسسته هم صادق باشد. چون در حالت گسسته، تابع چگالی یک سری δ است و نرمال که پیوسته است، نمی‌تواند آن را تقریب بزند (اگر چه نرمال، پوش مقدار وزنهٔ این δها را تقریب می‌زند.

به عنوان مثال قضیهٔ دموار- لاپلاس حالت خاصی از CLT است.

مثال: $\mathbf{x}_i$ها متغیرهای تصادفی .i.i.d و دارای توزیع برنولی هستند، یعنی:

$$P\{\mathbf{x}_i = 1\} = p \quad , \quad P\{\mathbf{x}_i = 0\} = q = 1 - p$$

حال اگر تعریف کنیم:

$$\mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i$$

$\mathbf{y}$ دارای توزیع دوجمله‌ای است، یعنی:

$$P\{\mathbf{y} = k\} = \binom{n}{k} p^k q^{n-k}$$

چون $\mathbf{x}_i$ها .i.i.d هستند و $\sigma_{\mathbf{x}}^2 = pq < +\infty$، پس طبق قضیهٔ لیندبرگ- لوی تابع توزیع $\mathbf{y}$ توسط نرمال قابل تقریب است (برای $n \to \infty$ به نرمال میل می‌کند). یعنی قضیهٔ دموار- لاپلاس را می‌توان حالت خاصی از CLT دانست.

اثبات قضیهٔ حد مرکزی در حالت .i.i.d بودن $\mathbf{x}_i$ها و محدود بودن تمامی گشتاورها:

$$\mathbf{y}=\sum_{i=1}^{n}\mathbf{x}_i \quad\Rightarrow\quad \eta_{\mathbf{y}}=n\eta \quad , \quad \sigma_{\mathbf{y}}^2=n\sigma^2$$

$$\mathbf{z}=\frac{\mathbf{y}-\eta_{\mathbf{y}}}{\sigma_{\mathbf{y}}}=\frac{\sum_{i=1}^{n}(\mathbf{x}_i-\eta)}{\sqrt{n}\,\sigma}=\frac{1}{\sqrt{n}}\overbrace{\sum_{i=1}^{n}(\underbrace{\frac{\mathbf{x}_i-\eta}{\sigma}}_{\mathbf{u}_i})}^{\mathbf{v}}$$

$$\Phi_{\mathbf{z}}(j\omega)=E(e^{j\omega\mathbf{z}})=E(e^{\frac{j\omega}{\sqrt{n}}\mathbf{v}})=\Phi_{\mathbf{v}}(\frac{j\omega}{\sqrt{n}})$$

اگر فرض کنیم $\mathbf{u}_i\triangleq\dfrac{\mathbf{x}_i-\eta}{\sigma}$، چون $\mathbf{v}=\sum_{i=1}^{n}\mathbf{u}_i$ و $\mathbf{u}_i$ها .i.i.d هستند، پس:

$$\Phi_{\mathbf{v}}(j\omega)=\Phi_{\mathbf{u}}^{n}(j\omega) \quad\Rightarrow\quad \Phi_{\mathbf{z}}(j\omega)=\Phi_{\mathbf{u}}^{n}(\frac{j\omega}{\sqrt{n}})$$

اما اگر تابع مشخصه را بسط تیلور دهیم، خواهیم داشت:

$$\Phi(j\omega) = 1 + j\omega\Phi'(0) + \frac{(j\omega)^2}{2}\Phi''(0) + \frac{(j\omega)^3}{3!}\Phi'''(0) + \cdots$$

که: $\Phi^{(n)}(0) = m_n$.

پس در مورد $\mathbf{u}$ داریم:

$$\Phi'(0) = \eta_{\mathbf{u}} = 0 \quad , \quad \Phi''(0) = \sigma_{\mathbf{u}}^2 = 1 \quad , \quad \Phi'''(0) = m_3(\mathbf{u}) = E\left[\left(\frac{\mathbf{x}_i - \eta}{\sigma}\right)^3\right]$$

در نتیجه:

$$\Phi_{\mathbf{u}}(j\omega) = 1 + \frac{(j\omega)^2}{2} + \frac{(j\omega)^3}{3!}\Phi'''(0) + \cdots \quad \Rightarrow \quad \Phi_{\mathbf{z}}(j\omega) = \left[1 + \frac{(j\omega)^2}{2n} + \frac{(j\omega)^3}{3!n^{\frac{3}{2}}}\Phi'''(0) + \cdots\right]^n$$

می‌دانیم که اگر $\lim_{n\to+\infty} a_n = b$ ، آنگاه:

$$\lim_{n\to+\infty}\left(1+\frac{a_n}{n}\right)^n = e^b$$

(یا با توجه به $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \mp \cdots$)

پس:

$$\Phi_{\mathbf{z}}(j\omega) = \left[1 + \frac{(j\omega)^2}{2n} + \frac{(j\omega)^3}{3!n^{\frac{3}{2}}}\Phi'''(0) + \cdots\right]^n$$

$$\to a_n = 1 + \frac{(j\omega)^2}{2} + \frac{(j\omega)^3}{3!n^{\frac{1}{2}}}\Phi'''(0) + \cdots \quad \Rightarrow \quad \lim_{n\to+\infty}\Phi_{\mathbf{z}}(j\omega) = e^{-\frac{\omega^2}{2}}$$

که تابع مشخصهٔ نرمال استاندارد است (قضیه‌ای داریم که اگر $\Phi_n \to \Phi^*$، آنگاه $F_n \to F^*$).

همگرایی دنبالهٔ متغیرهای تصادفی:

می‌دانیم که: $x_1, x_2, \ldots, x_n, \ldots$ یک دنباله از اعداد است و همگرایی این دنباله به عدد a، یعنی برای هر $\varepsilon > 0$، n_0ای وجود داشته باشد که:

$$|x_n - a| < \varepsilon : n > n_0$$

در این صورت می‌نویسیم:

$$\lim_{n \to +\infty} x_n = a$$

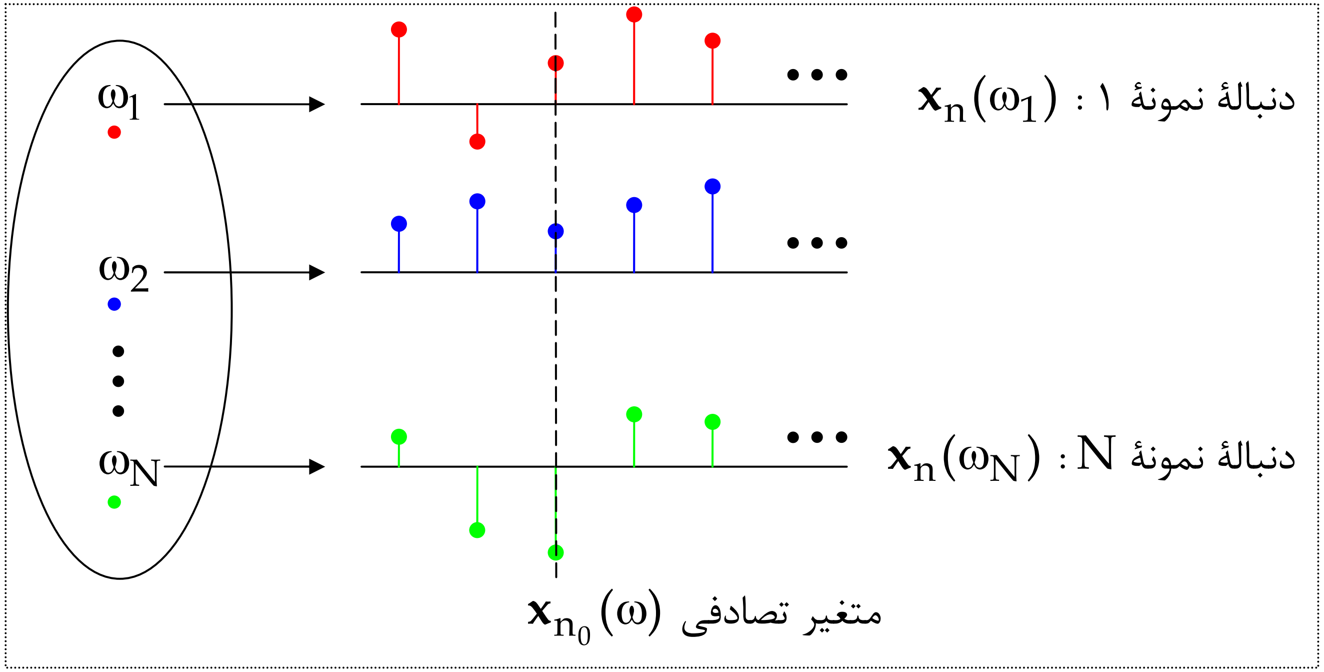
حال در مورد متغیرهای تصادفی مفهوم همگرایی به چه صورت خواهد بود؟

تعریف فرآیند تصادفی (Random Process or Stochastic Process):

دنبالهٔ نامحدود $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \ldots$ از متغیرهای تصادفی را فرآیند تصادفی گویند.

توجه دارید که هر $\mathbf{x}_i$ تابعی از نقاط فضای نمونه است، یعنی برای هر نقطهٔ ω_i یک دنبالهٔ عددی به صورت زیر داریم:

$$\mathbf{x}_1(\omega_i), \mathbf{x}_2(\omega_i), \ldots, \mathbf{x}_n(\omega_i), \ldots$$



توجه داشته باشید که:

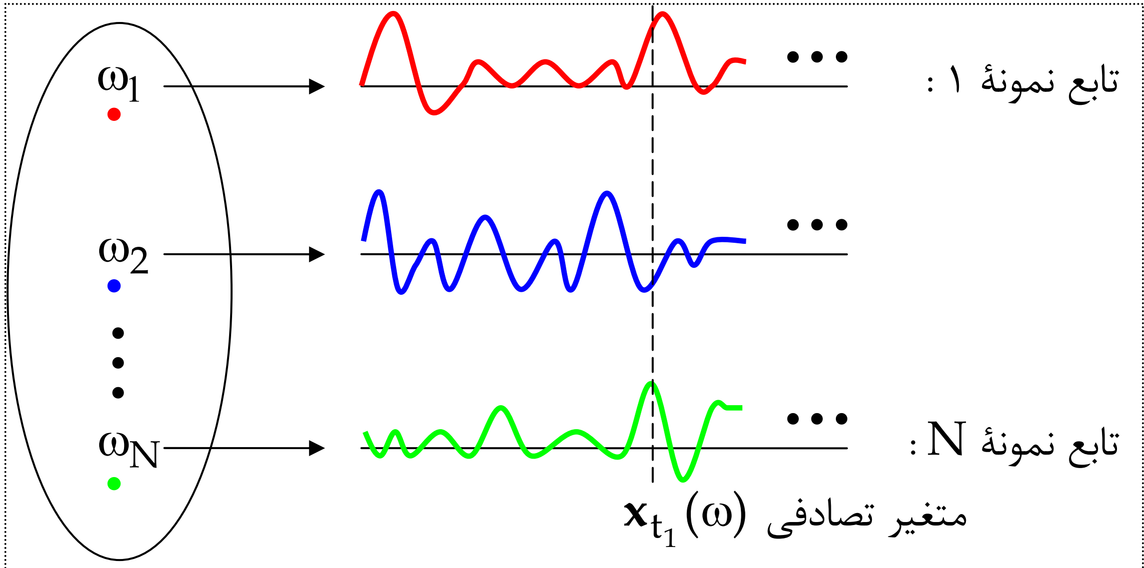
فرآیند تصادفی : $\mathbf{x}_n(\omega)$

دنبالهٔ عددی : $\mathbf{x}_n(\omega_i)$

متغیر تصادفی : $\mathbf{x}_{n_0}(\omega)$

یک عدد : $\mathbf{x}_{n_0}(\omega_i)$

آنچه گفتیم فرآیند تصادفی گسسته بود (سیگنال گسسته در زمان). به طور مشابه می‌توانیم فرآیند تصادفی پیوستهٔ $\mathbf{x}_t(\omega)$ را تعریف کنیم.



مثال: شیر یا خط می‌اندزیم. اگر شیر آمد، سیگنال سینوسی و اگر خط آمد سیگنال مربعی می‌فرستیم، یعنی به احتمال $\frac{1}{2}$ سینوسی و به احتمال $\frac{1}{2}$ مربعی.

مثال: برای نویز روی مقاومت، مقدار ولتاژ نویز در یک لحظه، متغیر تصادفی است. ولی شکل موج روی آن در یک مدت زمانی، فرآیند تصادفی است.

در مورد همگرایی دنبالهٔ متغیرهای تصادفی ممکن است $\mathbf{x}_n(\omega)$ برای بعضی ωها همگرا باشد و برای بعضی دیگر نباشد. یک تعریف برای همگرایی دنبالهٔ متغیر تصادفی می‌تواند این باشد که کلیهٔ دنباله‌های نمونه، یعنی $\mathbf{x}_n(\omega)$ برای هر ω همگرا باشد (everywhere). ولی در بسیاری موارد ممکن است این طور نباشد و همگرایی ضعیف‌تری داشته باشیم.

۱. همگرایی تقریباً همه‌جا (Almost Everywhere):

به این نوع همگرایی، همگرایی .a.e، همگرایی به احتمال یک و نیز همگرایی Almost Sure گفته می‌شود.

$\mathbf{x}_n(\omega)$ ممکن است برای بعضی ωها حد داشته و برای بعضی دیگر حد نداشته باشد. اگر مجموعهٔ ωهایی‌که برای آنها $\mathbf{x}_n(\omega)$ حد ندارد احتمالش صفر باشد، یعنی:

$$P\{\omega : \mathbf{x}_n(\omega) \xrightarrow[n\to+\infty]{} \mathbf{x}\} = 1$$

گوییم $\mathbf{x}_n$ دارای همگرایی .a.e است. در حالت کلی $\mathbf{x}$ می‌تواند تابع ω باشد (یعنی عدد ثابتی نباشد، بلکه متغیر تصادفی باشد).

در این صورت می‌نویسیم:

$$\mathbf{x}_n \xrightarrow[n\to+\infty]{\text{a.e.}} \mathbf{x}$$

۲. همگرایی در احتمال:

یعنی برای هر $\varepsilon > 0$ داشته باشیم:

$$P\{|\mathbf{x}_n - \mathbf{x}| \geq \varepsilon\} \xrightarrow[n\to+\infty]{} 0$$

اگر $\mathbf{x} = c$، یعنی عددی ثابت باشد، این به آن معناست که توزیع $\mathbf{x}_n$ با افزایش n حول c متمرکزتر می‌شود. در این صورت می‌نویسیم:

$$\mathbf{x}_n \xrightarrow[n\to+\infty]{P} \mathbf{x}$$

یا:

$$\underset{n\to+\infty}{P\lim}\, \mathbf{x}_n = \mathbf{x}$$

روشن است که همگرایی به احتمال یک، همگرایی در احتمال را نتیجه می‌دهد.

۳. همگرایی به مفهوم $m.s.$ (Mean Square) (همگرایی در گشتاور دوم):

برای هر n، $\mathbf{x}_n - \mathbf{x}$ یک متغیر تصادفی است و گشتاور دوم آن $E\left[(\mathbf{x}_n - \mathbf{x})^2\right]$ می‌باشد. پس $E\left[(\mathbf{x}_n - \mathbf{x})^2\right]$ یک دنبالهٔ عددی است.

می‌گوییم $\mathbf{x}_n$ به $\mathbf{x}$ به مفهوم $m.s.$ میل می‌کند و می‌نویسیم: $\mathbf{x}_n \xrightarrow[n \to +\infty]{m.s.} \mathbf{x}$، هرگاه:

$$E\left[(\mathbf{x}_n - \mathbf{x})^2\right] \xrightarrow[n \to +\infty]{} 0$$

همگرایی به مفهوم $m.s.$، همگرایی در احتمال را نتیجه می‌دهد.

زیرا با توجه به نامساوی Bienayme (به ازای $a = 2$) داریم:

$$P\{|\mathbf{x}_n - \mathbf{x}| > \varepsilon\} \le \frac{E\left[(\mathbf{x}_n - \mathbf{x})^2\right]}{\varepsilon^2}$$

پس وقتی $E\left[(\mathbf{x}_n - \mathbf{x})^2\right]$ به سمت صفر میل می‌کند، این احتمال هم به صفر میل می‌کند.

پس همگرایی $m.s.$ و همگرایی به احتمال یک $(a.e.)$ قویتر از همگرایی در احتمالند.

۴. همگرایی در تابع توزیع:

x_n را همگرا به $\mathbf{x}$ در تابع توزیع گویند، هرگاه برای هر x (هر xای که $F_{\mathbf{x}}(x)$ پیوسته باشد) داشته باشیم:

$$F_{\mathbf{x}_n}(x) \xrightarrow[n\to+\infty]{} F_{\mathbf{x}}(x)$$

این ضعیف‌ترین نوع همگرایی است. CLT نمونه‌ای از همگرایی در تابع توزیع بود.

قانون اعداد بزرگ:

در فصل ۳ دیدیم که اگر در یک آزمایش تصادفی $P(A)=p$ باشد و در n بار تکرار آزمایش، k بار واقعهٔ A اتفاق افتد، احتمال اینکه $\frac{k}{n}$ بین $p-\varepsilon$ و $p+\varepsilon$ باشد، وقتی $n\to+\infty$ به سمت یک می‌رود، یعنی:

برای هر $\varepsilon>0$ داده شده : $P\{|\frac{k}{n}-p|\le\varepsilon\}\to 1$

حال در حالت کلی‌تر داریم:

قانون ضعیف اعداد بزرگ (Weak Law of Large Numbers):

اگر $\mathbf{x}_i$ها نمونه‌های تصادفی از جامعه‌ای با میانگین η و .i.i.d باشند و $\overline{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$، میانگین نمونه باشد، داریم:

$$\underset{n\to+\infty}{\mathrm{P}\lim}\,\overline{\mathbf{x}}_n = \eta \qquad \text{یا} \qquad \overline{\mathbf{x}}_n \xrightarrow[n\to+\infty]{P} \eta$$

یعنی:

$$P\{|\overline{\mathbf{x}}_n - \eta| > \varepsilon\} \xrightarrow[n\to+\infty]{} 0$$

اثبات: می‌دانیم که:

$$E(\overline{\mathbf{x}}) = \eta \quad , \quad \sigma^2_{\overline{\mathbf{x}}} = \frac{\sigma^2}{n}$$

لذا با توجه به قضیهٔ چبیشف خواهیم داشت:

$$P\{|\overline{\mathbf{x}} - \eta| > \varepsilon\} < \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0$$

یا می‌توانیم بگوییم که:

$$E\left[(\overline{\mathbf{x}}_n-\eta)^2\right]=\sigma^2_{\overline{\mathbf{x}}_n}=\frac{\sigma^2}{n}\xrightarrow[n\to+\infty]{}0$$

پس:

$$\overline{\mathbf{x}}_n\xrightarrow[n\to+\infty]{\text{m.s.}}\eta$$

در نتیجه:

$$\overline{\mathbf{x}}_n\xrightarrow[n\to+\infty]{\text{P}}\eta$$

قانون قوی اعداد بزرگ (Strong Law of Large Numbers):

اگر $\mathbf{x}_i$ها متغیرهای تصادفی .i.i.d باشند و $\overline{\mathbf{x}}_n=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ باشد، $\overline{\mathbf{x}}_n$ با احتمال یک به η میل می‌کند.

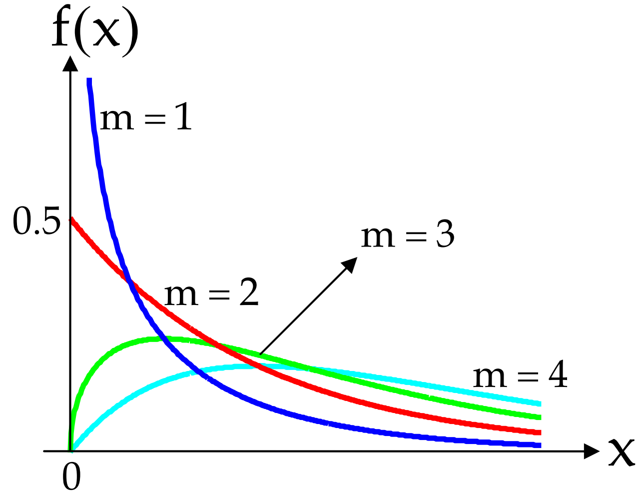
$$\overline{\mathbf{x}}_n\xrightarrow[n\to+\infty]{\text{a.e.}}\eta$$

توابع توزیع متداول در آمار:

توزیع χ^2: این توزیع کاربرد زیادی در آمار دارد. بنا بر تعریف، توزیع χ^2 با m درجه آزادی برابر است با:

$$\mathbf{x} \sim \chi^2(m) = \mathrm{Gamma}(r = \frac{m}{2}, \lambda = \frac{1}{2})$$

$$f_{\mathbf{x}}(x) = \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} u(x)$$



به ازای $m = 2$، این توزیع همان توزیع نمایی می‌شود.

داریم:

$$\Gamma(\frac{m}{2}) = \begin{cases} (k-1)! & m = 2k \\ (k-\frac{1}{2})(k-\frac{3}{2})\cdots\frac{1}{2}\sqrt{\pi} & m = 2k+1 \end{cases}$$ (چون $\Gamma(\frac{1}{2}) = \sqrt{\pi}$)

قبلاً گشتاور مرتبهٔ nام توزیع گاما را به دست آورده بودیم:

$$m_n = \frac{\Gamma(n+r)}{\Gamma(r)\lambda^n}$$

پس برای توزیع χ^2 داریم:

$$E(\mathbf{x}^n) = \frac{\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\,(\frac{1}{2})^n}$$

پس خواهیم داشت:

$$E(\mathbf{x}) = m \quad , \quad E(\mathbf{x}^2) = m(m+2) \quad \Rightarrow \quad var(\mathbf{x}) = 2m$$

رابطهٔ $m_n = \dfrac{\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\,(\frac{1}{2})^n}$ حتی برای n منفی صادق است (مشروط بر اینکه $n+\frac{m}{2} > 0$ باشد). مثلاً داریم:

$$E(\frac{1}{\mathbf{x}}) = \frac{\Gamma(\frac{m}{2}-1)}{2\Gamma(\frac{m}{2})} = \frac{1}{2(\frac{m}{2}-1)} = \frac{1}{m-2} : m > 2$$

تابع مشخصهٔ توزیع گاما را نیز قبلاً به دست آورده بودیم (در تمرین سری ۵):

$$\Phi(j\omega)=\left(\frac{\lambda}{\lambda-j\omega}\right)^{r}$$

پس برای توزیع χ^2 داریم:

$$\Phi_{\mathbf{x}}(j\omega)=\left(\frac{\frac{1}{2}}{\frac{1}{2}-j\omega}\right)^{\frac{m}{2}}=\frac{1}{(1-2j\omega)^{\frac{m}{2}}}$$

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مستقل باشند و $\mathbf{x}_i \sim \chi^2(k_i)$ بوده و $\mathbf{z}=\sum_{i=1}^{n}\mathbf{x}_i$ باشد، در این صورت داریم:

$$\mathbf{z}\sim\chi^2\left(\sum_{i=1}^{n}k_i\right)$$

در تمرین سری هشتم این قضیه را برای توزیع گاما دیدیم. پس برای توزیع χ^2 نیز که حالت خاصی از توزیع گاما است نیز صادق می‌باشد $(\lambda=\frac{1}{2})$.

قضیهٔ خاصیت اساسی توزیع χ^2**:** اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مستقل و نرمال استاندارد باشند و $\mathbf{Q}=\sum_{i=1}^{n}\mathbf{x}_i^2$ باشد، در این صورت داریم:

$$\mathbf{Q}\sim\chi^2(n)$$

اثبات:

$$\begin{cases}\mathbf{y}_i=\mathbf{x}_i^2\\ \mathbf{x}_i\sim N(0,1)\end{cases}\quad\Rightarrow\quad f_{\mathbf{y}_i}(y)=\frac{1}{\sqrt{y}}e^{-\frac{y}{2}}:y>0\quad\Rightarrow\quad\mathbf{y}_i\sim\chi^2(1)$$

پس طبق قضیهٔ قبل چون: $\mathbf{Q}=\sum_{i=1}^{n}\mathbf{x}_i^2=\sum_{i=1}^{n}\mathbf{y}_i$، داریم:

$$\mathbf{Q}\sim\chi^2(n)$$

(به طور مستقیم هم می‌توان نشان داد، ولی قدری مشکل است. اثبات در کتاب فرآیند Papoulis، صفحهٔ ۲۵۱.)

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مستقل و نرمال استاندارد باشند، فرم درجهٔ دوم $\mathbf{Q} = \underline{\mathbf{x}}^T A \underline{\mathbf{x}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\mathbf{x}_i\mathbf{x}_j$ (که در آن A یک ماتریس متقارن $n \times n$ است) دارای توزیع $\chi^2(r)$ خواهد بود، اگر و تنها اگر r مقدار ویژهٔ ماتریس A برابر یک و سایر $n-r$ مقدار ویژهٔ آن برابر صفر باشند.

اثبات: از جبر خطی می‌دانیم که ماتریس A را می‌توان به صورت زیر نوشت:

$$A = U\Lambda U^{-1} = U\Lambda U^T$$

که در آن:

$$U = \begin{bmatrix} \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_n \end{bmatrix} \quad , \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

زیرا از تعریف بردار ویژه و مقدار ویژه داریم:

$$A\underline{u}_i = \lambda_i \underline{u}_i$$

$$\Rightarrow \left[A\underline{u}_1 \quad A\underline{u}_2 \quad \cdots \quad A\underline{u}_n \right] = \left[\lambda_1 \underline{u}_1 \quad \lambda_2 \underline{u}_2 \quad \cdots \quad \lambda_n \underline{u}_n \right]$$

$$\Rightarrow A\left[\underline{u}_1 \quad \underline{u}_2 \quad \cdots \quad \underline{u}_n \right] = \left[\underline{u}_1 \quad \underline{u}_2 \quad \cdots \quad \underline{u}_n \right] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

$$\Rightarrow AU = U\Lambda \quad \Rightarrow \quad A = U\Lambda U^{-1}$$

البته اگر A متقارن باشد، U ماتریس Unitary خواهد بود، یعنی: $U^{-1} = U^T$. زیرا:

$$\underline{u}_i \cdot \underline{u}_j^T = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

به عبارت دیگر:

$$\Lambda = U^T A U$$

اگر تعریف کنیم:

$$\underline{\mathbf{z}} = U^T \underline{\mathbf{x}} \quad \Rightarrow \quad \underline{\mathbf{x}} = U \underline{\mathbf{z}}$$

$$E(\underline{\mathbf{z}}) = U^T E(\underline{\mathbf{x}}) = 0$$

$$C_{\underline{\mathbf{z}}} = U^T C_{\underline{\mathbf{x}}} U = U^T I U = U^T U = I$$

یعنی $\mathbf{z}_i$ها نرمال استاندارد و مستقل هستند.

$$\mathbf{Q} = \underline{\mathbf{x}}^T A \underline{\mathbf{x}} = \underline{\mathbf{z}}^T U^T A U \underline{\mathbf{z}} = \underline{\mathbf{z}}^T \Lambda \underline{\mathbf{z}} = \sum_{i=1}^{n} \lambda_i \mathbf{z}_i^2$$

پس $\mathbf{Q}$ توزیع $\chi^2(r)$ خواهد داشت، اگر و تنها اگر λ_iها صرفاً صفر یا یک باشند (r تعداد مقادیر ویژهٔ غیرصفر $(\lambda = 1)$ است).

قضیه: اگر برای سه فرم درجه دوم $\mathbf{Q}_1$، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ داشته باشیم: $\mathbf{Q}_3 = \mathbf{Q}_2 - \mathbf{Q}_1$ که: $\mathbf{Q}_1 \sim \chi^2(r)$، $\mathbf{Q}_2 \sim \chi^2(n)$ و $\mathbf{Q}_3 \geq 0$ باشند، آنگاه $\mathbf{Q}_1$ و $\mathbf{Q}_3$ مستقلند و $\mathbf{Q}_3 \sim \chi^2(n-r)$ است.

از کاربردهای مهم توزیع χ^2 در بررسی واریانس نمونه از توزیع نرمال است.

واریانس نمونه:

قبلاً دیدیم که میانگین نمونه برابر است با:

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \rightarrow \quad E(\overline{\mathbf{x}}) = \eta \quad , \quad \sigma_{\overline{\mathbf{x}}}^2 = \frac{\sigma^2}{n}$$

طبق تعریف، واریانس نمونه برابر است با:

$$\mathcal{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^2$$

در تمرین (سری هشتم) نشان دادید که:

$$E\left[(\mathbf{x}_i - \overline{\mathbf{x}})^2\right] = \frac{n-1}{n}\sigma^2$$

پس:

$$E(\mathcal{s}^2) = \frac{1}{n}\cdot n(\frac{n-1}{n}\sigma^2) = \frac{n-1}{n}\sigma^2$$

یعنی $\mathscr{s}^2$ تخمینی نااریب از σ^2 نیست. البته اگر η را می‌دانستیم، تخمینی نااریب از σ^2 می‌شد، زیرا:

$$\mathscr{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \eta)^2 \quad \rightarrow \quad E(\mathscr{s}^2) = \frac{1}{n} \cdot n\sigma^2 = \sigma^2$$

ولی وقتی η و σ^2 هیچکدام معلوم نیستند، معمولاً از تخمین زیر استفاده می‌شود که نااریب است (اگر چه بعداً خواهیم دید که واریانس این تخمین قدری بیشتر است):

$$\mathbf{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^2$$: واریانس نمونه

(بعضاً این تعریف را با $\mathbf{s}_{n-1}^2$ و تعریف اولی را با $\mathbf{s}_n^2$ نشان می‌دهند. برای n بزرگ، این دو تعریف یکسان می‌شوند. برای n کوچک معمولاً از تعریف دومی استفاده می‌شود.)

با این تعریف خواهیم داشت:

$$E(\mathbf{s}^2) = \frac{1}{n-1} \cdot n \cdot \frac{n-1}{n}\sigma^2 = \sigma^2$$

قضیه: میانگین نمونه و واریانس نمونهٔ یک متغیر تصادفی نرمال، مستقل از هم هستند و داریم:

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2} = \frac{n\mathscr{s}^2}{\sigma^2} \sim \chi^2(n-1)$$

اثبات:

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{\mathbf{x}_i - \overline{\mathbf{x}}}{\sigma}\right)^2 = \underbrace{\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i - \eta}{\sigma}\right)^2}_{\mathbf{Q}_2} - \underbrace{\left(\frac{\overline{\mathbf{x}} - \eta}{\frac{\sigma}{\sqrt{n}}}\right)^2}_{\mathbf{Q}_1} \quad \rightarrow \quad \mathbf{Q}_2 \sim \chi^2(n)\ ,\ \mathbf{Q}_1 \sim \chi^2(1)$$

پس با توجه به قضیهٔ قبلی، $\mathbf{Q}_3 = \mathbf{Q}_2 - \mathbf{Q}_1$ دارای توزیع $\chi^2(n-1)$ است، یعنی:

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2} \sim \chi^2(n-1)$$

و نیز $\mathbf{Q}_1$ و $\mathbf{Q}_3$ مستقلند. در نتیجه میانگین نمونه ($\overline{\mathbf{x}}$) و واریانس نمونه ($\mathscr{s}^2$ یا $\mathbf{s}^2$) مستقلند.

حال که فهمیدیم $\frac{(n-1)\mathbf{s}^2}{\sigma^2}=\frac{n\mathcal{s}^2}{\sigma^2}\sim\chi^2(n-1)$ است و می‌دانیم که واریانس متغیر تصادفی $\chi^2(m)$، $2m$ است، داریم:

$$\operatorname{var}(\mathbf{s}^2)=\left(\frac{\sigma^2}{n-1}\right)^2\cdot 2(n-1)=\frac{2\sigma^4}{n-1}$$

$$\operatorname{var}(\mathcal{s}^2)=\left(\frac{\sigma^2}{n}\right)^2\cdot 2(n-1)=\frac{2(n-1)}{n^2}\sigma^4$$

(یعنی $\mathbf{s}^2$ اگر چه ناریب است، ولی واریانس بیشتری دارد. همچنین ملاحظه می‌کنید که وقتی $n\to+\infty$، واریانس به سمت صفر می‌رود.)

توزیع χ^2 غیر مرکزی (Noncentral χ^2 Distribution):

اگر $\mathbf{x}_i$ها به جای اینکه $N(0,1)$ باشند، $N(\eta_i,1)$ بوده و مستقل نیز باشند، $\mathbf{Q}=\sum_{i=1}^{n}\mathbf{x}_i^2$ دیگر توزیع χ^2 ندارد، بلکه توزیع آن χ^2 غیرمرکزی است:

$\mathbf{Q}\sim\chi^2(n,e) \quad \rightarrow \quad$ n : درجهٔ آزادی / e : خروج از مرکز (Eccentricity)

$$e=\sum_{i=1}^{n}\eta_i^2 \quad , \quad E(\mathbf{Q})=n+e$$

همچنین اگر $\mathbf{x}_i\sim N(\eta_i,1)$، $\mathbf{Q}=\underline{\mathbf{x}}^T A\underline{\mathbf{x}}$ و ماتریس A دارای r مقدار ویژهٔ یک و $n-r$ مقدار ویژهٔ صفر باشد، آنگاه $\mathbf{Q}\sim\chi^2(r,e)$ خواهد بود که:

$$e=\eta_{\underline{\mathbf{x}}}^T A\,\eta_{\underline{\mathbf{x}}}$$

اثبات در کتاب، صفحات ۲۲۷ و ۲۲۸.

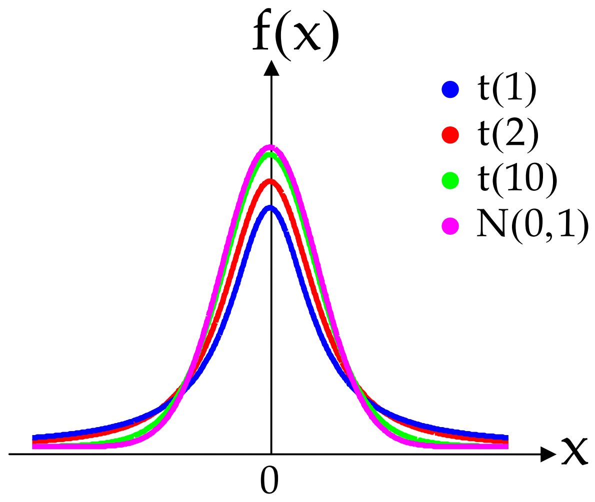
توزیع t (Student t) :

Gosset در سال ۱۹۰۸ تحت نام مستعار Student این توزیع را معرفی کرد:

$\mathbf{x} \sim t(m) \quad \rightarrow \quad$ m : درجهٔ آزادی

$$f_{\mathbf{x}}(x) = \gamma \frac{1}{(1+\frac{x^2}{m})^{\frac{m+1}{2}}} \quad \rightarrow \quad \gamma = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{\pi m}\,\Gamma(\frac{m}{2})}$$

این توزیع برای $m = 1$ همان توزیع کوشی است و برای $n \rightarrow +\infty$ می‌توان نشان داد که توزیع نرمال می‌شود. یعنی هر چه m کوچکتر باشد، دُم pdf درازتر است.



برای توزیع $t(n)$، m_k برای $k \geq n$ وجود ندارد، چون $E(|\mathbf{x}|^k)$ نامحدود می‌شود. مثلاً دیده بودیم که در توزیع کوشی، $E(\mathbf{x})$ نداریم. پس داریم:

$$E(\mathbf{x}) = 0 : m > 1$$

$$E(\mathbf{x}^2) = \operatorname{var}(\mathbf{x}) = \frac{m}{m-2} : m > 2$$

قضیهٔ خاصیت اساسی توزیع t: اگر $\mathbf{z} \sim N(0,1)$ و $\mathbf{w} \sim \chi^2(m)$ بوده و $\mathbf{z}$ و $\mathbf{w}$ مستقل باشند، در این صورت $\mathbf{x} = \dfrac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{m}}}$ دارای توزیع $t(m)$ خواهد بود، یعنی:

$$\mathbf{x} \sim t(m)$$

برای اثبات، مثلاً به روش متغیر کمکی می‌توانیم از روی $f_{\mathbf{zw}}$، $f_{\mathbf{x}}$ را به دست آوریم ($\mathbf{x} = \dfrac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{m}}}$ و $\mathbf{y} = \mathbf{w}$).

اثبات در کتاب، صفحهٔ ۲۳۳.

ضمناً اگر $\mathbf{z} \sim N(e,1)$ باشد، $\mathbf{x} \sim t(m,e)$ خواهد بود، یعنی توزیع t غیرمرکزی (Noncentral t Distribution) با m درجه آزادی و خروج از مرکز e خواهد داشت.

با توجه به قضیهٔ فوق برای توزیع t داریم:

$$E(\mathbf{x}^2) = m \cdot E(\mathbf{z}^2)E(\frac{1}{\mathbf{w}}) = m \cdot 1 \cdot \frac{1}{m-2} = \frac{m}{m-2}$$

همچنین داریم:

$$t(m) \xrightarrow[m \to +\infty]{} N(0,1)$$

زیرا اگر $\mathbf{x} = \dfrac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{m}}}$ باشد که $\mathbf{z} \sim N(0,1)$ و $\mathbf{w} \sim \chi^2(m)$ هستند، در مورد متغیر تصادفی $\mathbf{w}$ که دارای توزیع $\chi^2(m)$ است، خواهیم داشت:

$$\begin{cases} E(\mathbf{w}) = m \\ var(\mathbf{w}) = 2m \end{cases} \Rightarrow \begin{cases} E(\frac{\mathbf{w}}{m}) = 1 \\ var(\frac{\mathbf{w}}{m}) = \frac{2}{m} \end{cases}$$

پس: $\frac{\mathbf{w}}{m} \xrightarrow[m \to +\infty]{m.s.} 1$، یعنی: $\mathbf{x} \to \mathbf{z}$.

برای mهای کمتر $(m > 20)$ داریم:

$$\mathbf{x} \simeq \mathbf{z}\sqrt{\frac{m}{m-2}}$$

زیرا: $E(\mathbf{x}^2) = \frac{m}{m-2}$، ولی: $E(\mathbf{z}^2) = 1$.

کاربرد توزیع t: اگر $\bar{\mathbf{x}}$ و $\mathbf{s}^2$ میانگین و واریانس نمونه از متغیرهای تصادفی نرمال $N(\eta,\sigma)$ باشند و داشته باشیم:

$$\mathbf{T} = \frac{\bar{\mathbf{x}} - \eta}{\frac{\mathbf{s}}{\sqrt{n}}}$$

در این صورت داریم:

$$\mathbf{T} \sim t(n-1)$$

(در حالی که می‌دانیم $\frac{\bar{\mathbf{x}} - \eta}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد است.)

اثبات: با توجه به قضیه‌ای که داشتیم، $\mathbf{z}=\dfrac{\bar{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}}$ و $\mathbf{w}=(n-1)\dfrac{\mathbf{s}^2}{\sigma^2}$ مستقلند و $\mathbf{z}\sim N(0,1)$ و $\mathbf{w}\sim\chi^2(n-1)$ هستند.

پس طبق قضیهٔ بالا خواهیم داشت:

$$\mathbf{T}=\frac{\bar{\mathbf{x}}-\eta}{\frac{\mathbf{s}}{\sqrt{n}}}=\frac{\dfrac{\bar{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\dfrac{(n-1)\frac{\mathbf{s}^2}{\sigma^2}}{n-1}}}=\frac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{n-1}}}\sim t(n-1)$$

توجه کنید که مقدار $\mathbf{T}$ و توزیع آن ربطی به σ^2 پروسهٔ نرمال مورد نمونه‌گیری ندارد.

نکتهٔ دیگر اینکه برای $n\to+\infty$ که تخمین $\mathbf{s}$ از σ بهتر و بهتر می‌شود، توزیع متغیر تصادفی فوق به نرمال استاندارد میل می‌کند، یعنی:

$$t(n)\xrightarrow[n\to+\infty]{}N(0,1)$$

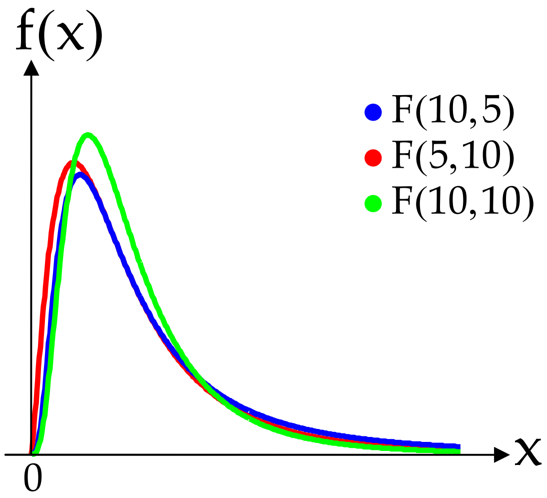
توزیع F (توزیع فیشر):

این توزیع در مسائل مختلف تست فرضیه در آمار ظاهر می‌شود.

گوییم **x** دارای توزیع F با درجه‌های آزادی k و m است: $\mathbf{x} \sim F(k,m)$، هرگاه:

$$f_{\mathbf{x}}(x) = \gamma \frac{x^{\frac{k}{2}-1}}{(1+\frac{kx}{m})^{\frac{k+m}{2}}} : x > 0 \quad \rightarrow \quad \gamma = \frac{\Gamma(\frac{k+m}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{m}{2})}\left(\frac{k}{m}\right)^{\frac{k}{2}}$$

(ترتیب k و m مهم است.)



قضیهٔ خاصیت اساسی توزیع F: اگر $\mathbf{z} \sim \chi^2(k)$ و $\mathbf{w} \sim \chi^2(r)$ بوده و مستقل باشند و داشته باشیم:

$$\mathbf{x} = \frac{\frac{\mathbf{z}}{k}}{\frac{\mathbf{w}}{r}}$$

آنگاه:

$$\mathbf{x} \sim F(k,r)$$

مشابه توزیع t، مثلاً با روش متغیر تصادفی کمکی می‌توان $f_{\mathbf{x}}$ را به دست آورد و نشان داد که فرم فوق‌الذکر را دارد.

اثبات در کتاب، صفحات ۲۲۴ و ۲۲۵.

ضمناً اگر $\mathbf{z} \sim \chi^2(k,e)$ و $\mathbf{w} \sim \chi^2(r)$ باشند، در این صورت: $\mathbf{x} \sim F(k,r,e)$، یعنی دارای توزیع F غیرمرکزی با درجه‌های آزادی k و r و خروج از مرکز e خواهد بود.

به توجه به قضیهٔ فوق برای توزیع F داریم:

$$E(\mathbf{x}) = \frac{r}{k} E(\mathbf{z}) E(\frac{1}{\mathbf{w}}) = \frac{r}{k} \cdot k \cdot \frac{1}{r-2} = \frac{r}{r-2} : r > 2$$

نتیجهٔ ۱:

$$\mathbf{x} \sim F(k,r) \quad \Rightarrow \quad \frac{1}{\mathbf{x}} \sim F(r,k)$$

نتیجهٔ ۲:

$$\mathbf{x} \sim t(m) \quad \Rightarrow \quad \mathbf{x}^2 \sim F(1,m)$$

زیرا:

$$\begin{cases} \mathbf{x} \sim \dfrac{\mathbf{z}}{\sqrt{\dfrac{\mathbf{w}}{m}}} \\ \\ \mathbf{z} \sim N(0,1) \\ \mathbf{w} \sim \chi^2(m) \end{cases} \Rightarrow \begin{cases} \mathbf{x}^2 \sim \dfrac{\mathbf{z}^2}{\dfrac{\mathbf{w}}{m}} \\ \\ \mathbf{z}^2 \sim \chi^2(1) \\ \mathbf{w} \sim \chi^2(m) \end{cases} \Rightarrow \quad \mathbf{x}^2 \sim F(1,m)$$

نتیجهٔ ۳:

$$\mathbf{x} \sim F(k,r) \quad \Rightarrow \quad \underset{r\to+\infty}{k\mathbf{x}} \sim \chi^2(k)$$

زیرا:

$$\begin{cases} k\mathbf{x} = \dfrac{\mathbf{z}}{\frac{\mathbf{w}}{r}} \\ \mathbf{z} \sim \chi^2(k) \\ \mathbf{w} \sim \chi^2(r) \Rightarrow \dfrac{\mathbf{w}}{r} \xrightarrow[r\to+\infty]{} 1 \end{cases} \quad \Rightarrow \quad k\mathbf{x} \xrightarrow[r\to+\infty]{} \mathbf{z}$$