5장. 표본분포의 근사

5.1 중심극한정리

이항 확률의 정규 근사(부록 p.502~505):

$$\sum_{x:\, a \le \frac{x-np}{\sqrt{np(1-p)}} \le b} \binom{n}{x} p^x (1-p)^{n-x} \sim \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz, \quad n \to \infty$$

[Key]

(1)(스털링의 공식) $\Gamma(m+1) \sim m^{m+1/2} e^{-m} \sqrt{2\pi}, \quad m \to \infty$

(2)(이항 확률밀도의 근사) $a \le \dfrac{x-np}{\sqrt{npq}} \le b$ 인 x에 대하여

$$\binom{n}{x} p^x (1-p)^{n-x} \sim \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-np}{\sqrt{npq}}\right)^2\right\} \frac{1}{\sqrt{npq}}, \quad n \to \infty \qquad (q \equiv 1-p)$$

(3)(누적 이항 확률의 근사)

$$\sum_{x:\, a \le \frac{x-np}{\sqrt{npq}} \le b} \binom{n}{x} p^x (1-p)^{n-x} \sim \frac{1}{\sqrt{npq}} \sum_{x:\, a \le \frac{x-np}{\sqrt{npq}} \le b} \phi\left(\frac{x-np}{\sqrt{npq}}\right)$$

$$\sim \frac{b-a}{N-1} \sum_{j=1}^{N} \phi(z_j)$$

$$\sim \int_a^b \phi(z) dz, \qquad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

정리 5.1.1:중심극한정리(中心極限整理 central limit theorem)

확률변수 $X_1, \cdots, X_n$이 서로 독립이고 동일한 분포를 따르며 $\mathrm{Var}(X_1)$이 양의 실수일 때

$$E(X_1) = \mu, \mathrm{Var}(X_1) = \sigma^2 \ (0 < \sigma < +\infty)$$

이라고 하면, 표준정규분포 $N(0,1)$을 따르는 확률변수 Z에 대하여 다음이 성립한다.

$$\lim_{n \to \infty} P\left(\frac{(X_1 + \cdots + X_n)/n - \mu}{\sigma/\sqrt{n}} \le x\right) = P(Z \le x) \quad \forall x: -\infty < x < +\infty$$

[증명의 개요] $\qquad mgf_{\sqrt{n}(\overline{X}_n - \mu)/\sigma}(t) = [mgf_{(X_1 - \mu)/\sigma}(t/\sqrt{n})]^n$

여기에서 $m(s) = mgf_{(X_1-\mu)/\sigma}(s)$라고 하면 적률생성함수의 성질과 테일러 정리로부터

$$m\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{1}{2}\frac{t^2}{n} + R_{n,t}, \qquad \lim_{n \to \infty} n R_{n,t} = 0$$

$$\log mgf_{\sqrt{n}(\overline{X}_n - \mu)/\sigma}(t) = n \log\left(1 + \frac{1}{2}\frac{t^2}{n} + R_{n,t}\right), \quad \lim_{n \to \infty} n R_{n,t} = 0$$

$$= n\left\{\frac{1}{2}\frac{t^2}{n} + r_{n,t}\right\}, \quad \lim_{n \to \infty} n r_{n,t} = 0$$

$$\therefore \lim_{n \to \infty} mgf_{\sqrt{n}(\overline{X}_n - \mu)/\sigma}(t) = \exp(t^2/2)$$

예 5.1.1

(a)(포아송분포의 정규근사)

서로 독립이고 포아송분포 Poisson(λ)를 따르는 $X_1,\cdots,X_n$에 대하여

$$E(X_1)=Var(X_1)=\lambda \ (0<\lambda<+\infty)$$

이므로

$$\lim_{n\to\infty} P\left(\frac{(X_1+\cdots+X_n)/n-\lambda}{\sqrt{\lambda/n}}\le x\right)=P(Z\le x),\quad Z\sim N(0,1)$$

한편 포아송분포의 성질로부터 $X_1+\cdots+X_n\sim Poisson(n\lambda)$이므로

$$\lim_{n\to\infty}\sum_{k:\frac{k-n\lambda}{\sqrt{n\lambda}}\le x}\frac{e^{-n\lambda}(n\lambda)^k}{k!}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}dz$$

따라서 <u>$Y_n\sim Poisson(n)$이고 n이 충분히 크면</u> 표준정규분포의 누적분포함수 Φ를 이용하여 다음과 같이 근사계산을 할 수 있다.[1]

$$P(a<Y_n\le b)\simeq\Phi\left(\frac{b-n}{\sqrt{n}}\right)-\Phi\left(\frac{a-n}{\sqrt{n}}\right)$$

(b)(감마분포 또는 카이제곱분포의 정규근사)

서로 독립이고 감마분포 Gamma(α,β)를 따르는 $X_1,\cdots,X_n$에 대하여

$$E(X_1)=\alpha\beta,\ Var(X_1)=\alpha\beta^2\ (\alpha>0,\beta>0)$$

이므로

$$\lim_{n\to\infty} P\left(\frac{(X_1+\cdots+X_n)/n-\alpha\beta}{\sqrt{\alpha\beta^2/n}}\le x\right)=P(Z\le x),\quad Z\sim N(0,1)$$

한편 $Y_n\sim\chi^2(n)$이면

$$Y_n\overset{d}{\equiv}X_1+\cdots X_n,\ X_i\overset{iid}{\sim}\chi^2(1)=Gamma(1/2,2)\ (i=1,\cdots,n)$$

이므로

$$\lim_{n\to\infty} P\left(\frac{Y_n-n}{\sqrt{2n}}\le x\right)=P(Z\le x),\quad Z\sim N(0,1)$$

따라서 <u>$Y_n\sim\chi^2(n)$이고 n이 충분히 크면</u> (a)에서와 마찬가지로

$$P(a<Y_n\le b)\simeq\Phi\left(\frac{b-n}{\sqrt{2n}}\right)-\Phi\left(\frac{a-n}{\sqrt{2n}}\right)$$

---

1) 이러한 근사계산은 중심극한정리에서 정규분포로의 수렴이 <u>균등수렴</u>(연습문제 5.9 참조)이기 때문에 가능한 것이다.

정리 5.1.2: 다차원 경우의 중심극한정리

다차원 확률변수 $\mathrm{X}_1=(\mathrm{X}_{11},\cdots,\mathrm{X}_{1k})^t,\cdots,\mathrm{X}_n=(\mathrm{X}_{n1},\cdots,\mathrm{X}_{nk})^t$이 서로 독립이고 동일한 분포를 따르며 분산행렬 $\mathrm{Var}(\mathrm{X}_1)$이 존재할 때

$$\mathrm{E}(\mathrm{X}_1)=\mu, \mathrm{Var}(\mathrm{X}_1)=\Sigma$$

라고 하면, $\mathrm{N}_k(0,\Sigma)$를 따르는 $\mathrm{Z}=(\mathrm{Z}_1,\cdots,\mathrm{Z}_k)^t$ 에 대하여 다음이 성립한다.[2)]

$$\lim_{n\to\infty}\mathrm{P}\left\{\sqrt{n}((\mathrm{X}_{1j}+\cdots+\mathrm{X}_{nj})/n-\mu_j)\le x_j,\ j=1,\cdots,k\right\}=\mathrm{P}(\mathrm{Z}_1\le x_1,\cdots,\mathrm{Z}_k\le x_k)\ \forall x_j$$

예 5.1.2 다항분포의 다변량 정규근사:

다항분포의 대의적 정의로부터 $\mathrm{X}_n=(\mathrm{X}_{n1}\cdots,\mathrm{X}_{nk})^t\sim\mathrm{Multi}(n,(p_1,p_2,\cdots,p_k)^t)$ 이면

$$\mathrm{X}_n\overset{d}{\equiv}\mathrm{Z}_1+\cdots+\mathrm{Z}_n,\ \mathrm{Z}_i=(\mathrm{Z}_{i1},\mathrm{Z}_{i2},\cdots,\mathrm{Z}_{ik})^t\overset{iid}{\sim}\mathrm{Multi}(1,(p_1,p_2,\cdots,p_k)^t)\ (i=1,\cdots,n)$$

한편 $p_1,\cdots,p_k$ 를 대각원소로 하는 대각행렬을 $D(p_j)$라고 하면 정리 3.2.2로부터

$$\mathrm{E}(\mathrm{X}_1)=p=(p_1,\cdots,p_k)^t,\ \ \mathrm{Var}(\mathrm{X}_1)=D(p_j)-pp^t,$$

따라서 중심극한정리로부터

$$\lim_{n\to\infty}\mathrm{P}\left\{(\mathrm{X}_n-np)/\sqrt{n}\le x\right\}=\mathrm{P}(\mathrm{Z}\le x),\ \mathrm{Z}\sim\mathrm{N}_k(0,D(p_j)-pp^t)$$

이 때 분산행렬 $D(p_j)-pp^t$은 특이행렬로서 역행렬을 갖지 않는 점에 유의하여야 한다.

---

2) 다차원 확률변수에 관한 이러한 결합확률을 간략히 다음과 같이 나타내기도 한다.

$$\lim_{n\to\infty}\mathrm{P}\left\{\sqrt{n}((\mathrm{X}_1+\cdots+\mathrm{X}_n)/n-\mu)\le x\right\}=\mathrm{P}(\mathrm{Z}\le x),\ \forall x\in R^k,\ \mathrm{Z}\sim\mathrm{N}_k(0,\Sigma)$$

5.2 극한분포와 확률적 수렴

예 5.2.1(확률분포를 근사할 때 모든 점에서 수렴할 것을 요구하지 않아도 되는 예)

동전을 던져서 앞면이 나오면 0부터 $1+n^{-1}$사이의 수를 랜덤하게 선택하여 보여주고, 뒷면이 나오면 $1+n^{-1}$을 보여주는 실험에서 보게 될 숫자를 $\mathrm{X}_n\,(n=1,2,\cdots)$이라고 하자.

확률변수 $\mathrm{X}_n\,(n=1,2,\cdots)$의 누적분포함수가 다음과 같은 경우에 분포의 근사를 생각해보자.

$$cdf_{\mathrm{X}_n}(x)=\mathrm{P}(\mathrm{X}_n \le x)=\begin{cases}0, & x<0\\ \dfrac{1}{2}\dfrac{1}{1+n^{-1}}x, & 0\le x<1+n^{-1}\\ 1, & x\ge 1+n^{-1}\end{cases}$$

이러한 누적분포함수들의 극한으로 정의되는 함수는

$$\mathrm{G}(x)=\lim_{n\to\infty} cdf_{\mathrm{X}_n}(x)=\begin{cases}0,\ x<0\\ \dfrac{1}{2}x,\ 0\le x\le 1\\ 1,\ x>1\end{cases}$$

로서, $x=1$의 오른쪽에서 연속이 아닌 함수이다. 따라서 정리 1.5.1로부터 이 함수는 누적분포함수가 아니다.

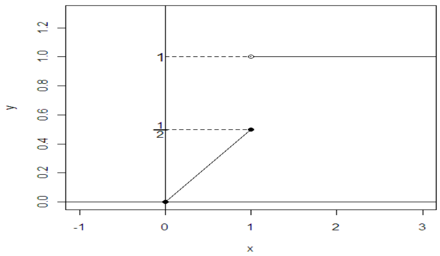

그림 1.5.1 누적분포함수의 극한이 연속이 아닌 경우

예 5.2.1에서 극한 함수가 연속이 아닌 점 $x=1$의 오른쪽에서 연속이도록 수정한 함수를

$$\mathrm{F}(x)=\begin{cases}0,\ x<0\\ \dfrac{1}{2}x,\ 0\le x<1\\ 1,\ x\ge 1\end{cases}$$

라고 하면 함수 $\mathrm{F}(x)$는 누적분포함수이고, 예 5.2.1의 확률변수들 $\mathrm{X}_\mathrm{n}\,(\mathrm{n}=1,2,\cdots)$에 대하여

$$\lim_{n\to\infty} cdf_{\mathrm{X}_n}(x)=\mathrm{F}(x)\quad \forall\, x : x\ne 1$$

임을 알 수 있다. 이 경우에 동전을 던져서 앞면이 나오면 0부터 1사이의 수를 랜덤하게 선택하여 보여주고, 뒷면이 나오면 1을 보여주는 실험에서 보게 될 숫자를 X라고 하면 함수 함수 $\mathrm{F}(x)$는 확률변수 X의 누적분포함수이다. 따라서

$$\lim_{n\to\infty} cdf_{\mathrm{X}_n}(x)=cdf_{\mathrm{X}}(x)\quad \forall\, x : x\ne 1$$

이고 $\mathrm{X}_n\,(n=1,2,\cdots)$의 분포가 X의 분포로 근사된다고 할 수 있을 것이다.

극한분포(極限分布 limiting distribution):
확률변수의 열 $X_n (n=1,2,\cdots)$과 Z에 대하여

$$\lim_{n\to\infty} P(X_n \le x) = P(Z \le x)$$

가 Z의 누적분포함수 cdf_Z가 연속인 모든 점 x에서 성립할 때, Z의 분포를 X_n 분포들의 극한분포 또는 점근분포(漸近分布 asymptotic distribution)라고 하며 기호로는

$$X_n \underset{n\to\infty}{\overset{d}{\to}} Z$$

로 나타낸다. 즉 cdf_Z가 연속인 점들의 집합을 $\mathrm{Conti}(cdf_Z)$라고 하면

$$X_n \underset{n\to\infty}{\overset{d}{\to}} Z \iff \lim_{n\to\infty} cdf_{X_n}(x) = cdf_Z(x) \quad \forall x : x \in \mathrm{Conti}(cdf_Z)$$

예 5.2.2 이항분포의 포아송근사:
이항분포 $B(n, \lambda/n)\,(0<\lambda/n<1)$을 따르는 확률변수 X_n과 포아송분포 $\mathrm{Poisson}(\lambda)$를 따르는 확률변수 X에 대하여 다음이 성립하는 것은 3장 4절에서 소개되었다.

$$P(X_n = k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \underset{n\to\infty}{\to} \frac{1}{k!}\lambda^k e^{-\lambda} = P(X=k),\ k=1,2,\cdots$$

한편 이들 확률변수들은 0 또는 자연수의 값만을 가지므로 모든 점 x에서

$$cdf_{X_n}(x) = \sum_{k:0\le k\le x} P(X_n = k) \underset{n\to\infty}{\to} \sum_{k:0\le k\le x} P(X=k) = cdf_X(x)$$

임을 알 수 있다. 따라서 X의 분포 $\mathrm{Poisson}(\lambda)$는 X_n의 분포 $B(n,\lambda/n)$의 극한분포이고 이를 기호로 다음과 같이 나타내기도 한다.

$$B(n,\lambda/n) \approx \mathrm{Poisson}(\lambda),\ n\to\infty$$

예 5.2.3
균등분포 $U(0,1)$에서의 랜덤표본 n개에 기초한 순서통계량을 $U_{(1)} < \cdots < U_{(n)}$이라고 할 때

$$P\{n(1-U_{(n)}) \le x\} = 1 - P\left\{U_{(n)} < 1 - \frac{x}{n}\right\}$$

이고

$$P\left\{U_{(n)} < 1-\frac{x}{n}\right\} = \begin{cases} 0,\ 1-x/n<0 \text{ 즉 } x>n \\ (1-\frac{x}{n})^n,\ 0\le 1-x/n<1 \text{ 즉 } 0<x\le n \\ 1,\ 1\le 1-x/n \text{ 즉 } x\le 0 \end{cases}$$

$$\lim_{n\to\infty} P\left\{U_{(n)} < 1-\frac{x}{n}\right\} = \begin{cases} e^{-x}, & 0<x<+\infty \\ 1, & x\le 0 \end{cases}$$

이므로 $n(1-U_{(n)})$의 극한분포를 다음과 같이 구할 수 있다.

$$\lim_{n\to\infty} P\{n(1-U_{(n)}) \le x\} = \begin{cases} 1-e^{-x}, & x\ge 0 \\ 0, & x<0 \end{cases}$$

$$\therefore\ n(1-U_{(n)}) \underset{n\to\infty}{\overset{d}{\to}} Z,\quad Z \sim \mathrm{Exp}(1)$$

정리 5.2.1: 극한분포가 상수의 분포인 경우

확률변수의 열 $X_n\,(n=1,2,\cdots)$의 극한분포가 상수 c의 분포일 조건은 다음과 같다.

$$X_n \underset{n\to\infty}{\overset{d}{\to}} X,\ P(X=c)=1 \Leftrightarrow \lim_{n\to\infty} P(|X_n-c|\ge\epsilon)=0 \quad \forall\,\epsilon>0$$

[Key] $P(X=c)=1$인 경우에 그 누적분포함수는

$$cdf_X(x)=\begin{cases}0, & x<c\\ 1, & x\ge c\end{cases}$$

이므로 cdf_X는 $x=c$이외의 모든 점에서 연속인 함수이다.

확률수렴(確率收斂 converge in probability): $X_n \underset{n\to\infty}{\overset{P}{\to}} c$ 또는 $\mathrm{p}\lim\limits_{n\to\infty} X_n = c$

$$\mathrm{p}\lim_{n\to\infty} X_n = c \quad\Leftrightarrow\quad \lim_{n\to\infty} P(|X_n-c|\ge\epsilon)=0 \quad \forall\,\epsilon>0$$

정리 5.2.2: 대수의 법칙(大數의 法則 law of large numbers)

확률변수 $X_1,\cdots,X_n$이 서로 독립이고 동일한 분포를 따르며 $E(X_1)$이 실수로 정의되면[3)]

$$\mathrm{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = E(X_1)$$

[증명] 여기에서는 추가적인 조건으로서 $Var(X_1)<+\infty$인 경우만을 다루기로 한다.

체비셰프의 부등식을 표본평균 $\overline{X_n}=(X_1+\cdots+X_n)/n$에 적용하면, 임의의 양수 ϵ에 대하여

$$0 \le P(|\overline{X_n}-E(X_1)|\ge\epsilon) = P(|\overline{X_n}-E(\overline{X_n})|\ge\epsilon) \le Var(\overline{X_n})/\epsilon^2 = Var(X_1)/n\epsilon^2$$

$$\therefore \lim_{n\to\infty} P(|\overline{X_n}-E(X_1)|\ge\epsilon)=0$$

랜덤표본 $X_1,\cdots,X_n$을 관측할 때 관측값이 A에 속하는 상대도수는

$$\frac{1}{n}\sum_{i=1}^{n} I_A(X_i)$$

로 주어진다. 이 상대도수에 대수의 법칙을 적용하면 다음이 성립하는 것을 알 수 있다.

$$\mathrm{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} I_A(X_i) = E(I_A(X_1)) = P(X_1\in A)$$

즉 시행 회수가 많아짐에 따라 상대도수가 확률에 한없이 가까이 가는 것을 뜻한다.

다차원 경우의 확률수렴:

다차원 확률변수 $X_n=(X_{n1},\cdots,X_{nk})^t$과 상수의 벡터 $c=(c_1,\cdots,c_k)^t$에 대하여

$$\|\, X_n - c\,\| = \sqrt{(X_{n1}-c_1)^2+\cdots+(X_{nk}-c_k)^2}$$

라고 할 때

$$\lim_{n\to\infty} P(\|X_n-c\|\ge\epsilon)=0 \quad \forall\,\epsilon>0$$

이면 $X_n=(X_{n1},\cdots,X_{nk})^t$이 $c=(c_1,\cdots,c_k)^t$로 확률수렴한다고 한다.

3) 기댓값의 정의에 따라 이는 $E(|X_1|)<+\infty$임을 뜻한다.

정리 5.2.3

다차원 확률변수 $X_n = (X_{n1}, \cdots, X_{nk})^t$이 상수의 벡터 $c = (c_1, \cdots, c_k)^t$로 확률수렴하는 것 즉 $\operatorname{p\lim}_{n\to\infty} X_n = c$ 와 다음의 조건들은 같은 뜻이다.

(a) $\lim_{n\to\infty} P(\max_{1\le i\le k} |X_{ni} - c_i| \ge \epsilon) = 0$

(b) $\operatorname{p\lim}_{n\to\infty} X_{n1} = c_1, \cdots, \operatorname{p\lim}_{n\to\infty} X_{nk} = c_k$

[증명의 Key] (a) $\max_{1\le i\le k} |X_{ni} - c_i| \le \| X_n - c \| \le k \max_{1\le i\le k} |X_{ni} - c_i|$

$$\therefore P(\max_{1\le i\le k} |X_{ni} - c_i| \ge \epsilon) \le P(\| X_n - c \| \ge \epsilon) \le P(\max_{1\le i\le k} |X_{ni} - c_i| \ge \epsilon/k)$$

(b) $$(|X_{ni} - c_i| \ge \epsilon) \subseteq (\max_{1\le i\le k} |X_{ni} - c_i| \ge \epsilon) = \bigcup_{i=1}^{k} (|X_{ni} - c_i| \ge \epsilon) \ (i = 1, \cdots, k)$$

$$\max_{1\le i\le k} P(|X_{ni} - c_i| \ge \epsilon) \le P(\max_{1\le i\le k} |X_{ni} - c_i| \ge \epsilon) \le \sum_{i=1}^{k} P(|X_{ni} - c_i| \ge \epsilon)$$

정리 5.2.4 :다차원 경우에 대수의 법칙

다차원 확률변수 $X_1 = (X_{11}, \cdots, X_{1k})^t, \cdots, X_n = (X_{n1}, \cdots, X_{nk})^t$이 서로 독립이고 동일한 분포를 따르며 $E(X_1) = (E(X_{11}), \cdots, E(X_{1k}))^t$이 정의될 수 있으면[4)]

$$\operatorname{p\lim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = E(X_1)$$

[증명] 일차원 경우에 대수의 법칙과 정리 5.2.3으로부터 이 정리는 명백히 성립한다.

정리 5.2.5: 연속 함수와 확률수렴

다차원 확률변수 $X_n = (X_{n1}, \cdots, X_{nk})^t$이 상수의 벡터 $c = (c_1, \cdots, c_k)^t$로 확률수렴하고 실수값 함수 g가 c 에서 연속이면 $g(X_n)$이 $g(c)$로 확률수렴한다. 즉 함수 g가 c 에서 연속일 때

$$\operatorname{p\lim}_{n\to\infty} X_n = c \text{ 이면 } \operatorname{p\lim}_{n\to\infty} g(X_n) = g(c)$$

[증명] 함수 g가 c 에서 연속이므로, 다음이 성립한다.

$$\forall \epsilon > 0,\ \exists \delta > 0 : (\|x - c\| < \delta \Rightarrow |g(x) - g(c)| < \epsilon)$$

따라서 임의의 양수 ϵ에 대하여 이러한 양수 δ를 선택하면

$$(\| X_n - c \| < \delta) \subseteq (|g(X_n) - g(c)| < \epsilon)$$

$$\therefore (|g(X_n) - g(c)| \ge \epsilon) \subseteq (\| X_n - c \| \ge \delta)$$

$$\therefore P(|g(X_n) - g(c)| \ge \epsilon) \le P(\| X_n - c \| \ge \delta)$$

그런데 $\operatorname{p\lim}_{n\to\infty} X_n = c$ 이므로

$$0 \le \lim_{n\to\infty} P(|g(X_n) - g(c)| \ge \epsilon) \le \lim_{n\to\infty} P(\| X_n - c \| \ge \delta) = 0$$

---

4) 기댓값의 정의에 따라 이는 $E(|X_{11}|) < +\infty, \cdots, E(|X_{1k}|) < +\infty$ 임을 뜻한다.

정리 5.2.6: 확률수렴과 사칙연산[5)]
확률변수 $X_n, Y_n\,(n=1,2,\cdots)$이 각각 실수 a,b로 확률수렴할 때 다음이 성립한다.
(a) $\mathrm{p}\lim_{n\to\infty}(X_n+Y_n)=\mathrm{p}\lim_{n\to\infty}X_n+\mathrm{p}\lim_{n\to\infty}Y_n=a+b$
(b) $\mathrm{p}\lim_{n\to\infty}(X_n-Y_n)=\mathrm{p}\lim_{n\to\infty}X_n-\mathrm{p}\lim_{n\to\infty}Y_n=a-b$
(c) $\mathrm{p}\lim_{n\to\infty}(X_n\times Y_n)=\mathrm{p}\lim_{n\to\infty}X_n\times\mathrm{p}\lim_{n\to\infty}Y_n=a\times b$
(d) $\mathrm{p}\lim_{n\to\infty}(X_n\div Y_n)=\mathrm{p}\lim_{n\to\infty}X_n\div\mathrm{p}\lim_{n\to\infty}Y_n=a\div b\quad(b\neq 0)$
[증명] 정리 5.2.3으로부터 이차원 확률변수 $(X_n,Y_n)^t$가 벡터$(a,b)^t$로 확률수렴한다. 즉

$$\mathrm{p}\lim_{n\to\infty}(X_n,Y_n)^t=(a,b)^t$$

따라서 정리 5.2.5에 의해 연속인 이변수 함수를 적용하면 확률수렴성이 보존된다.
 한편 사칙연산은 이변수 함수로서 그 정의역에서 연속인 함수이므로 이 정리가 성립한다. 구체적으로는 정리 5.2.5를 다음 함수들에 적용하면 이 정리의 결과를 얻게 된다.

$$g_1(x,y)=x+y,\ g_2(x,y)=x-y,\ g_3(x,y)=x\times y,\ g_4(x,y)=x\div y$$

예 5.2.4 표본적률(標本積率 sample moment) 벡터의 확률수렴:
랜덤표본 $X_1,\cdots,X_n$의 함수로서

$$\widehat{m_r}=\frac{1}{n}\sum_{i=1}^{n}X_i^r$$

을 r차 표본적률이라고 하며, 이는 모집단 적률인 $m_r=E(X_1^r)$의 추론에 사용된다. 한편 정리 1.6.2로부터 $E(|X_1^k|)<+\infty$이면 즉 k차 적률 $E(X_1^k)$가 실수로 정의될 수 있으면 더 낮은 차의 적률이 모두 실수로 정의될 수 있는 것을 알고 있다. 따라서 대수의 법칙으로부터 $E(|X_1^k|)<+\infty$이면

$$\mathrm{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(X_i,\cdots,X_i^k)^t=(E(X_1),\cdots,E(X_1^k))^t$$

 이로부터 표본적률의 벡터인

$$(\widehat{m_1},\cdots,\widehat{m_k})^t=\left(\frac{1}{n}\sum_{i=1}^{n}X_i,\cdots,\frac{1}{n}\sum_{i=1}^{n}X_i^k\right)^t=\frac{1}{n}\sum_{i=1}^{n}(X_i,\cdots,X_i^k)^t$$

은 표본크기가 커짐에 따라 그 추측 대상인 모집단의 적률벡터에 한없이 가까이 간다는 것을 알 수 있다.

5) 이 정리에서 확률변수 X_n이나 Y_n이 각각 상수 x_n이나 $y_n\,(n=1,2,\cdots)$인 경우에도 그 결과가 성립한다. 이는 X_n이나 Y_n이 하나의 값만 갖는 상수의 확률변수들인 경우로 생각할 수 있기 때문이다.

예 5.2.5 표본분산과 표본표준편차의 확률수렴:
랜덤표본 $X_1, X_2, \cdots, X_n (n \ge 2)$을 이용하여 모분산에 관한 추론에 사용되는 표본분산

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

은 다음과 같이 두 표본적률 $(\sum_{i=1}^{n} X_i/n, \sum_{i=1}^{n} X_i^2/n)^t$의 함수로 나타낼 수 있다.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n}X_i^2 - (\frac{1}{n}\sum_{i=1}^{n}X_i)^2\right\}$$

한편 예 5.2.4에서와 같이 대수의 법칙으로부터, $E(X_1^2) < +\infty$이면

$$\mathrm{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i^2 = E(X_1^2),\ \ \mathrm{p}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = E(X_1)$$

따라서 정리 5.2.6으로부터

$$\mathrm{p}\lim_{n\to\infty}S_n^2 = \mathrm{p}\lim_{n\to\infty}\frac{n}{n-1}\left\{\frac{1}{n}\sum_{i=1}^{n}X_i^2 - (\frac{1}{n}\sum_{i=1}^{n}X_i)^2\right\} = 1 \times \left\{E(X_1^2) - [E(X_1)]^2\right\}$$

그러므로 모분산 $\sigma^2 = \mathrm{Var}(X_1)$이 양의 실수일 때

$$\mathrm{p}\lim_{n\to\infty}S_n^2 = \sigma^2$$

또한 표본표준편차는 $S_n = \sqrt{S_n^2}$이고 제곱근함수는 연속함수이므로 정리 5.2.5로부터

$$\mathrm{p}\lim_{n\to\infty}S_n = \sqrt{\mathrm{p}\lim_{n\to\infty}S_n^2} = \sqrt{\sigma^2} = \sigma$$

정리 5.2.7
분산이 실수로 정의될 수 있는 확률변수 $X_n (n = 1, 2, \cdots)$에 대하여 다음이 성립한다.

$$\lim_{n\to\infty}\mathrm{Var}(X_n) = 0,\ \lim_{n\to\infty}E(X_n) = a \ \text{ 이면 }\ \mathrm{p}\lim_{n\to\infty}X_n = a$$

[증명] 정리 1.6.3의 마코프 부등식으로부터, 임의의 양수 ϵ에 대하여 다음이 성립한다.

$$P(|X_n - a| \ge \epsilon) \le E[(X_n - a)^2]/\epsilon^2$$

한편

$$E[(X_n - a)^2] = \mathrm{Var}(X_n) + \{E(X_n) - a\}^2$$

이므로, $\lim_{n\to\infty}\mathrm{Var}(X_n) = 0,\ \lim_{n\to\infty}E(X_n) = a$ 이면

$$0 \le \lim_{n\to\infty}P(|X_n - a| \ge \epsilon) \le \lim_{n\to\infty}E[(X_n - a)^2]/\epsilon^2 = 0$$

예 5.2.6
균등분포 $U(0,1)$에서의 랜덤표본 n개에 기초한 순서통계량을 $U_{(1)} < \cdots < U_{(n)}$이라고 할 때 $U_{(n)}$의 확률밀도함수가

$$pdf_{U_{(n)}}(x) = n x^{n-1} I_{(0,1)}(x)$$

이므로

$$E(U_{(n)}) = \frac{n}{n+1},\ \ \mathrm{Var}(U_{(n)}) = \frac{n}{n+2} - (\frac{n}{n+1})^2$$

따라서

$$\lim_{n\to\infty}\mathrm{Var}(U_{(n)}) = 0,\quad \lim_{n\to\infty}E(U_{(n)}) = 1$$

$$\therefore\ \mathrm{p}\lim_{n\to\infty}U_{(n)} = 1$$

예 5.2.7 표본분위수(標本分位數 sample quantile)[6]**의 확률수렴:**

모집단 분포가 연속형이고 그 누적분포함수 $\mathrm{F}(x)$가 순증가함수[7]일 때 누적확률이 α가 되는 값 $\mathrm{F}^{-1}(\alpha)\ (0<\alpha<1)$를 모집단의 α 분위수(分位數 quantile)라고 한다. 한편 랜덤표본 n 개에 기초한 순서통계량을 $\mathrm{X}_{(1)}<\cdots<\mathrm{X}_{(n)}$이라고 할 때

$$r_n \sim \alpha n,\ \text{즉}\ \lim_{n\to\infty} r_n/n=\alpha\ \ (0<\alpha<1)$$

를 만족하는 자연수 r_n 에 대하여 $\mathrm{X}_{(r_n)}$을 표본분위수라고 한다. 한편 정리 4.3.4로부터

$$\mathrm{X}_{(r_n)} \stackrel{d}{\equiv} h\left(\frac{1}{n}\mathrm{Z}_1+\cdots+\frac{1}{n-r_n+1}\mathrm{Z}_{r_n}\right),\quad \mathrm{Z}_i \stackrel{iid}{\sim} \mathrm{Exp}(1)\ (i=1,\cdots,n)$$

$$h(y)=\mathrm{F}^{-1}(1-e^{-y})\ (y>0)$$

한편 $\mathrm{Y}_n \equiv \frac{1}{n}\mathrm{Z}_1+\cdots+\frac{1}{n-r_n+1}\mathrm{Z}_{r_n}$이라고 하면

$$\mathrm{E}(\mathrm{Y}_n)=\frac{1}{n}+\cdots+\frac{1}{n-r_n+1},\quad \mathrm{Var}(\mathrm{Y}_n)=\frac{1}{n^2}+\cdots+\frac{1}{(n-r_n+1)^2}$$

$$\frac{1}{n}+\cdots+\frac{1}{n-r_n+1}=\frac{1}{n}\sum_{k=0}^{r_n-1}\frac{1}{1-k/n}\sim\int_0^\alpha\frac{1}{1-x}dx=-\log(1-\alpha)$$

$$\frac{1}{n^2}+\cdots+\frac{1}{(n-r_n+1)^2}=\frac{1}{n}\frac{1}{n}\sum_{k=0}^{r_n-1}\frac{1}{(1-k/n)^2}\sim\frac{1}{n}\int_0^\alpha\frac{1}{(1-x)^2}dx=\frac{1}{n}\frac{\alpha}{1-\alpha}$$

따라서

$$\lim_{n\to\infty}\mathrm{Var}(\mathrm{Y}_n)=\lim_{n\to\infty}\frac{1}{n}\frac{\alpha}{1-\alpha}=0,\qquad \lim_{n\to\infty}\mathrm{E}(\mathrm{Y}_n)=-\log(1-\alpha)$$

$$\therefore\ \mathrm{p}\lim_{n\to\infty}\mathrm{Y}_n=\mathrm{p}\lim_{n\to\infty}\left(\frac{1}{n}\mathrm{Z}_1+\cdots+\frac{1}{n-r_n+1}\mathrm{Z}_{r_n}\right)=-\log(1-\alpha)$$

그러므로 함수 $h(y)=\mathrm{F}^{-1}(1-e^{-y})\ (y>0)$가 연속인 경우에 정리 5.2.5로부터

$$\mathrm{p}\lim_{n\to\infty}h\left(\frac{1}{n}\mathrm{Z}_1+\cdots+\frac{1}{n-r_n+1}\mathrm{Z}_{r_n}\right)=h(-\log(1-\alpha))=\mathrm{F}^{-1}(\alpha)$$

이고, 정리 5.2.1의 필요조건으로부터 이러한 경우에

$$h\left(\frac{1}{n}\mathrm{Z}_1+\cdots+\frac{1}{n-r_n+1}\mathrm{Z}_{r_n}\right)\underset{n\to\infty}{\overset{d}{\rightarrow}}\mathrm{F}^{-1}(\alpha)$$

그런데 $\mathrm{X}_{(r_n)} \stackrel{d}{\equiv} h\left(\frac{1}{n}\mathrm{Z}_1+\cdots+\frac{1}{n-r_n+1}\mathrm{Z}_{r_n}\right)$이므로 이러한 경우에

$$\mathrm{X}_{(r_n)}\underset{n\to\infty}{\overset{d}{\rightarrow}}\mathrm{F}^{-1}(\alpha)$$

따라서 누적분포함수의 역함수 F^{-1}이 연속일 때 정리 5.2.1의 충분조건으로부터

$$\mathrm{p}\lim_{n\to\infty}\mathrm{X}_{(r_n)}=\mathrm{F}^{-1}(\alpha)$$

---

6) 표본분위수에 대한 이러한 정의는 표본 크기가 충분히 큰 경우에 사용하는 정의이며, 이는 예 4.1.2의 표본 중앙값의 정의와 다른 것을 알 수 있다. 이러한 정의는 확률수렴과 같은 극한 성질을 간략히 밝히기 위한 것이며, 일반적인 정의에 따른 표본분위수의 경우에도 모집단의 분위수로 확률수렴하는 것이 알려져 있다.

7) 확률적분변환에 관한 정리 4.3.3에서와 마찬가지로 누적분포함수가 순증가가 아닌 경우에는 F의 역변환을

$$\mathrm{F}^{-1}(u)=inf\{x:\mathrm{F}(x)\geq u\}\ (0\leq u\leq 1)$$

로 정의하고, 일반적으로는 이 역변환을 이용하여 분위수를 정의한다.