4. 감마함수와 감마분포

[정의 3.4-1] 감마함수 (Gamma function)

함수 $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx,\ n>0$를 감마함수라 한다.

[정리 3.4-1] 감마함수의 성질

(1) $\Gamma(1) = 1$

(2) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

(3) $\Gamma(n+1) = n!$, n이 자연수일 때

(4) $\Gamma(n+1) = n\Gamma(n),\ n>0$

(5) $\frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = \int_0^1 u^{m-1}(1-u)^{n-1}dx = Be(m,n)$

이 식은 베타함수(Beta function)의 정의이다.

(6) $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx = \sqrt{2\pi}$

증명 (1) $\Gamma(1) = \int_0^\infty e^{-x}dx = 1.$

(3), (4)
$$\begin{aligned}\Gamma(n+1) &= \int_0^\infty x^n e^{-x}dx \\ &= [-x^n e^{-x}]_0^\infty + n\int_0^\infty x^{n-1}e^{-x}dx \\ &= n\Gamma(n) \\ &\ \vdots \\ &= n!\end{aligned}$$

) 베타함수는 $Be(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx,\ m,\ n>0,\ x>0$으로 정의되므로 $x=1-y$로 치환하면

$$Be(m,n) = \int_0^1 (1-y)^{m-1}y^{n-1}dy = Be(n,m)$$

임을 알 수 있다. $x = \sin^2\theta$로 치환하면

$$Be(m,n) = 2\int_0^{\frac{\pi}{2}} \sin^{2m-1}\theta\cos^{2n-1}\theta\, d\theta$$

이므로 $\int_0^{\frac{\pi}{2}} \sin^p\theta\cos^q\theta\, d\theta = \frac{1}{2}Be\left(\frac{p+1}{2},\ \frac{q+1}{2}\right)$임을 알 수 있다.

감마함수의 정의에서 $x=y^2$로 치환하면 $dx = 2y\,dy$이므로 $\Gamma(n) = 2\int_0^\infty e^{-y^2}y^{2n-1}dy$가 된다. 따라서,

$$\Gamma(m)\Gamma(n) = 4\int_0^\infty\int_0^\infty e^{-(x^2+y^2)}x^{2m-1}y^{2n-1}dxdy$$

가 되며 극좌표 변환 ($x = r\cos\theta,\ y = r\sin\theta$)을 하면 $dxdy = r\,drd\theta$이므로

$$\begin{aligned}\Gamma(m)\Gamma(n) &= 4\int_0^{\frac{\pi}{2}}\int_0^\infty e^{-r^2}r^{2m+2n-1}\cos^{2m-1}\theta\sin^{2n-1}\theta\, drd\theta \\ &= \left(2\int_0^\infty e^{-r^2}r^{2m+2n-1}dr\right)\left(2\int_0^{\frac{\pi}{2}}\cos^{2m-1}\theta\sin^{2n-1}\theta\, d\theta\right) \\ &= \Gamma(m+n)\,Be(m,n)\end{aligned}$$

이다. 따라서 (5)는 증명되었다.

(2) $Be\left(\frac{1}{2},\ \frac{1}{2}\right) = 2\int_0^{\frac{\pi}{2}} d\theta = \pi = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma(1)}$이므로 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$이다.

(6)
$$\begin{aligned}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}dx &= 2\int_0^\infty e^{-\frac{x^2}{2}}dx = 2\int_0^\infty e^{-y}\frac{1}{\sqrt{2y}}dy \quad (y=x^2\text{으로 치환}) \\ &= \sqrt{2}\int_0^\infty e^{-y}y^{\frac{1}{2}-1}dy = \sqrt{2}\,\Gamma\left(\frac{1}{2}\right) = \sqrt{2\pi}.\end{aligned}$$

[정의 3.4-2] 확률변수 X를 pdf가

$$f(x) = \frac{\lambda^r x^{r-1}e^{-\lambda x}}{\Gamma(r)},\quad x>0,\quad r>0,\quad \lambda>0$$

로 주어질 때, X는 감마분포를 따른다. $X \sim \Gamma(r,\lambda)$

감마 분포의 밀도함수가 위의 정의와 같이 주어지는 근거를 살펴보자. 평균이 λ인 포아송 과정에서 처음 사건이 나타날 때까지의 대기시간은 지수 분포를 갖는다는 것을 앞 절에서 이미 다루었다. 확률변수 X를 r번째 사건이 일어날 때까지의 대기시간이라고 하면 X의 분포는 구간 $[0,x]$, $x>0$에서 사건의 수가 평균 λx인 포아송 분포를 갖게 되므로, 다음과 같이 표현된다.

$$\begin{aligned}F(x) &= P(X\le x) = 1 - P(X>x) \\ &= 1 - P([0,x]\text{에서 나타나는 사건이 } r\text{번 보다 적음}) \\ &= 1 - \sum_{k=0}^{r-1}\frac{(\lambda x)^k e^{-\lambda x}}{k!}\end{aligned}$$

$x>0$일 때, $F(x)$의 도함수 $F'(x)$는 확률변수 X의 밀도함수가 된다.

$$\begin{aligned}F'(x) &= \lambda e^{-\lambda x} - e^{-\lambda x}\sum_{k=1}^{r-1}\left[\frac{k(\lambda x)^{k-1}\lambda}{k!} - \frac{(\lambda x)^k\lambda}{k!}\right] \\ &= \lambda e^{-\lambda x} - e^{-\lambda x}\left[\lambda - \frac{\lambda(\lambda x)^{r-1}}{(r-1)!}\right] \\ &= \frac{\lambda(\lambda x)^{r-1}}{(r-1)!}e^{-\lambda x}\end{aligned}$$

물론 $x<0$이면 $F'(x) = 0$이다.

$f(x)$가 pdf의 조건을 갖추고 있는지 확인해 본다. $f(x)\ge 0$임은 분명하고

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_0^\infty \frac{\lambda^r x^{r-1}e^{-\lambda x}}{\Gamma(r)}dx = \int_0^\infty \frac{\lambda^r\left(\frac{y}{\lambda}\right)^{r-1}e^{-y}}{\Gamma(r)}\frac{1}{\lambda}dy\ (y=\lambda x\text{로 치환}) \\ &= \frac{1}{\Gamma(r)}\int_0^\infty y^{r-1}e^{-y}dy = 1\end{aligned}$$

이므로 pdf의 조건은 만족하게 된다.

[정리 3.4-2] 확률변수 X가 모수가 (r,λ)인 감마 분포를 갖는다면 X의 적률생성함수, 평균, 분산은 다음과 같다.

$$M(t) = \left(\frac{\lambda}{\lambda - t}\right)^r, \quad t < \lambda, \quad E(X) = \frac{r}{\lambda}, \quad Var(X) = \frac{r}{\lambda^2}.$$

증명

$$M(t) = \int_0^\infty e^{tx} \cdot \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} dx \ (y = (\lambda - t)x \text{로 치환})$$

$$= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty \frac{y^{r-1}}{(\lambda - t)^{r-1}} e^{-y} \left(\frac{1}{\lambda - t}\right) dy$$

$$= \frac{\lambda^r}{(\lambda - t)^r \Gamma(r)} \int_0^\infty y^{r-1} e^{-y} dy = \left(\frac{\lambda}{\lambda - t}\right)^r, \quad t < \lambda$$

$$M'(t) = \lambda^r(-r)(\lambda - t)^{-r-1}(-1) = r\lambda^r(\lambda - t)^{-r-1},$$

$$M''(t) = (r\lambda^r)(-r-1)(\lambda - t)^{-r-2}(-1) = (r\lambda^r)(r+1)(\lambda - t)^{-r-2}$$

이므로 평균과 분산은 아래와 같이 구해진다.

$$E(X) = M'(0) = \frac{r}{\lambda}, \quad E(X^2) = M''(0) = \frac{r(r+1)}{\lambda^2}, \quad Var(X) = \frac{r}{\lambda^2}.$$

[정리 3.4-3] 확률변수 $X_1, X_2, \cdots, X_r$이 서로 독립이고 모수가 λ인 지수분포를 따른다면, 확률변수 $Y = X_1 + X_2 + \cdots + X_r$은 모수가 (r, λ)인 감마분포를 따른다.

증명 확률변수 X_i의 적률생성함수는 $M_{X_i}(t) = \frac{\lambda}{\lambda - t}$ 이고 서로 독립이므로

$M_Y(t) = M_{X_1}(t) \cdot \cdots \cdot M_{X_r}(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$, $\lambda > 0$이다. 이는 모수가 (r, λ)인 감마 분포의 적률생성함수이다.

예 3.4-1 어떤 상점을 찾는 고객의 수는 1분에 평균 4명 $(\lambda = 4)$인 포아송 과정을 따른다고 한다. 다음 고객이 올 때까지 걸리는 시간을 T_1이라고 하면 T_1의 pdf는 모수가 4인 지수 분포를 따르며, pdf는 아래와 같다.

$$f_{T_1}(t) = 4e^{-4t}, \quad t > 0.$$

10번째 손님이 올 때까지 소요되는 시간을 T_{10}이라고 하면 T_{10}은 모수가 $(10, 4)$인 감마분포를 따르며 T_{10}의 pdf는 아래와 같다.

$$f_{T_{10}}(t) = \frac{4^{10} t^9 e^{-4t}}{9!}, \quad t > 0.$$

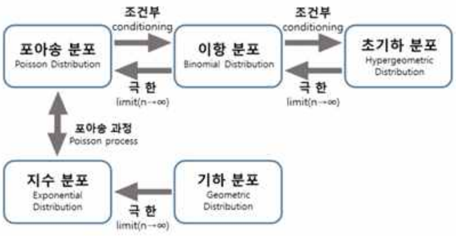
예 3.4-2 1시간에 평균 30명의 고객이 포아송 과정에 따라 상점에 들어온다. 즉, 1분당 평균 고객은 $\lambda = 1/2$명이다. 점원이 처음 두 고객이 들어올 때까지 5분 이상 기다릴 확률은 얼마인가?

풀이 확률변수 X를 두 번째 고객이 들어올 때까지의 대기시간(단위 : 분)이라고 하면 X는 모수가 $r = 2$, $\lambda = 1/2$인 감마분포를 갖는다. 따라서,

$$P(X > 5) = \int_5^\infty \frac{x^{2-1} e^{-x/2}}{\Gamma(2) 2^2} dx = \int_5^\infty \frac{x e^{-x/2}}{4} dx$$

$$= \frac{1}{4}[(-2)xe^{-x/2} - 4e^{-x/2}]_5^\infty = \frac{7}{2} e^{-5/2} = 0.287.$$

[참고] 서로 독립인 확률변수 $X_1, X_2, \cdots, X_k$가 각각 모수가 $(r_1, \lambda), \cdots, (r_k, \lambda)$인 감마분포를 따르면 확률변수 $Y = \sum_{i=1}^{k} X_i$의 분포는 모수가 $(r_1 + r_2 + \cdots + r_k, \lambda)$인 감마분포를 따른다.



5.베타분포와 카이제곱 분포; 화합물에 포함된 불순물의 비율, 특정공정 진행률

[정의 3.5-1] 확률변수 X를 pdf가

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},\quad \alpha,\ \beta>0,\quad 0<x<1$$

로 주어지는 경우에 X는 모수가 $(\alpha,\ \beta)$인 베타분포(Beta distribution)를 따른다

[정리 3.5-1] 확률변수 X는 모수가 (α,β)인 베타분포를 갖는다면 X의 평균과 분산

$$E(X)=\frac{\alpha}{\alpha+\beta},\quad Var(X)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

증명

$$\begin{aligned}E(X)&=\int_0^1 x\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx\\&=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)}\int_0^1\frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}x^{(\alpha+1)-1}(1-x)^{\beta-1}dx\\&=\frac{\alpha}{\alpha+\beta}.\end{aligned}$$

$$\begin{aligned}E(X^2)&=\int_0^1 x^2\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx\\&=\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha+\beta+2)\Gamma(\alpha)}\int_0^1\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+2)\Gamma(\beta)}x^{\alpha+1}(1-x)^{\beta-1}dx\\&=\frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)}\end{aligned}$$

$$Var(X)=E(X^2)-[E(X)]^2=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

[참고] ① $\int_0^a x^{m-1}(1-x)^{n-1}dx,\ 0<a<1$인 함수를 불완전 베타함수(Incomplete beta function)라고 한다.

② 확률변수 X_1이 모수가 $(\alpha_1,\ \lambda)$인 감마분포를 따르고, X_2가 모수 $(\alpha_2,\ \lambda)$인 감마분포를 따르면 변환된 확률변수 $Y_1=X_1+X_2$와 $Y_2=\frac{X_1}{X_1+X_2}$는 서로 독립이고 확률변수 Y_2는 베타분포를 갖는다.

[정의 3.5-2] 확률변수 X를 pdf가

$$f(x)=\frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}}x^{\frac{n}{2}-1}e^{-\frac{x}{2}},\quad x>0$$

일 때 X는 자유도가 n인 카이제곱분포(Chi-square distribution)를 따른다.

• 모수가 (r,λ)인 감마분포에서 $r=n/2,\ \lambda=1/2$인 경우가 카이제곱분포

[정리 3.5-2] 카이제곱분포를 갖는 확률변수 X의 적률생성함수, 평균, 분산

$$M(t)=(1-2t)^{-\frac{n}{2}},\ t<\frac{1}{2},\ E(X)=\frac{n}{2}\cdot 2=n,\ Var(X)=\frac{r}{\lambda^2}=\frac{\frac{n}{2}}{\left(\frac{1}{2}\right)^2}=2n.$$

카이제곱분포에서의 평균은 자유도와 같고 분산은 자유도의 2배

[참고] ① 확률변수 X가 모수 $\lambda=\frac{1}{2}$인 지수+분포를 따른다고 하면 X의 pdf는

$$f(x)=\frac{1}{2}e^{-\frac{x}{2}}=\frac{x^{\frac{2}{2}-1}e^{-\frac{x}{2}}}{\Gamma\left(\frac{2}{2}\right)2^{\frac{2}{2}}},$$ $0\le x<\infty$이므로 이는 자유도가 2인 카이제곱분포의 pdf와 같다.

② 모수가 λ인 지수분포는 모수가 $(1,\ \lambda)$인 감마분포와 같다. 더욱이 확률변수 X가 정규분포 $N(0,\ \sigma^2)$을 따른다면 X^2은 감마분포 $\Gamma\left(\frac{1}{2},\ \frac{1}{2\sigma^2}\right)$를 따른다고 알려져 있다.

6. 정규 분포

[정의 3.6-1] 확률변수 X를 pdf가

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x, \quad \mu < \infty, \quad 0 < \sigma < \infty$$

X

로 주어질 때 X는 모수가 (μ, σ^2)인 정규분포(Normal distribution)를 따른다.
$X \sim N(\mu, \sigma^2)$

• 정규분포의 밀도함수는 μ에 대하여 대칭이며 그래프의 모양은 종모양(Bell-shaped)을 하고 있다.

• $f(x) > 0$

$$I = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

라 두고 $z = \frac{(x-\mu)}{\sigma}$로 변수변환을 하면 $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$가 되며 $\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi}$이므로 $\int_{-\infty}^{\infty} f(x)dx = 1$이므로 밀도함수의 조건을 충족.

[정의 3.6-1] 확률변수 X가 $(0, 1)$인 정규분포를 표준정규분포 (Standard normal distribution)라고 하며, 밀도함수는

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

[정의 3.6-2] 표준정규 밀도함수의 분포함수

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$$

[참고] (표준정규 분포함수의 성질)
① $\Phi(\infty) = 1$
② $\Phi(0) = 0.5$
③ $\Phi(-x) = 1 - \Phi(x)$ (정규분포의 대칭성)

[예 3.6-1] 확률변수 Z의 분포가 표준정규분포 $N(0, 1)$일 때, 다음의 확률을 구해보자.

$$\begin{aligned} P[|Z-1| > 1.42] &= 1 - P[|Z-1| \le 1.42] \\ &= 1 - P[-0.42 \le Z \le 2.42] \\ &= 1 - \{\Phi(2.42) - \Phi(-0.42)\} \\ &= 1 - (0.9922 - 0.3372) = 0.3450. \end{aligned}$$

[정의 3.6-2] 확률변수 X가 정규분포 $N(\mu, \sigma^2)$을 따를 때, X의 적률생성함수, 평균, 분산

$$M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}, \quad E(X) = \mu, \quad Var(X) = \sigma^2.$$

• 정규확률분포의 확률밀도함수

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

• 정규확률분포의 적률생성함수

$$M(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma^2}\right)} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx)} dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x^2 - \mu - \sigma^2 t)^2} e^{\mu t + \frac{1}{2}\sigma^2 t^2} dx$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x^2 - \mu - \sigma^2 t)^2} dx \xrightarrow[\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x - (\mu + \sigma^2 t))^2} dx = 1]{\text{평균이 } \mu + \sigma^2 t\text{, 표준편차가 } \sigma \text{인 정규분포}} e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

• 정규확률분포의 평균: $R'(0) = E(X) = \mu$
• 정규확률분포의 분산: $R''(0) = Var(X) = \sigma^2$
• 정규확률분포의 누가적률

$R(t) = \mu t + \frac{1}{2}\sigma^2 t^2$, $R'(t) = \mu + \sigma^2 t$, $R''(t) = \sigma^2$

[정의 3.6-2] 확률변수 X가 정규분포 $N(\mu, \sigma^2)$을 따르면 다음이 성립한다.

(1) 변환된 확률변수 $Z = \frac{X-\mu}{\sigma}$는 표준정규분포 $N(0, 1)$을 따른다.

(2) $P[a < X < b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$.

증명 (1) $P[Z \le z] = P\left[\frac{X-\mu}{\sigma} \le z\right] = P[X \le \mu + \sigma z]$

$$= \int_{-\infty}^{\mu+\sigma z} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

⟨ $w = \frac{x-\mu}{\sigma}$ 로 치환하면,⟩

$$= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw = \Phi(z).$$

(2) $P[a \le X \le b] = P\left[\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$.

[참고] 정규분포 $N(\mu, \sigma^2)$을 따르는 확률변수 X를 $Z = \frac{X-\mu}{\sigma}$로 변환하는 경우에 확률변수 Z는 확률변수 X를 표준화 (Standardize)한다고 한다.

[예제 3.6-2] 확률변수 X의 pdf가

$$f(x) = \frac{1}{\sqrt{32\pi}} \exp\left[-\frac{(x+7)^2}{32}\right], \quad -\infty < x < \infty$$

로 주어지면 X의 분포는 $N(-7, 16)$이다. 즉, 확률변수 X는 평균이 -7이고, 분산이 16인 정규분포를 가지며 적률생성함수는 $M(t) = e^{-7t+8t^2}$이다.

[예 3.6-3] 확률변수 Z의 분포가 표준정규분포 $N(0, 1)$을 따를 때 다음 확률을 구해보자.

$$P[0 \le Z \le 2] = \Phi(2) - \Phi(0) = 0.9772 - 0.5000 = 0.4772,$$

$$\begin{aligned} P[-1.65 \le Z \le 0.70] &= \Phi(0.70) - \Phi(-1.65) \\ &= \Phi(0.70) - [1 - \Phi(1.65)] \\ &= 0.7580 - 1 + 0.9505 = 0.7085. \end{aligned}$$

[예 3.6-4] 확률변수 Z의 분포가 표준정규분포 $N(0, 1)$을 따를 때, $P[0 \le Z \le a] = 0.4147$, $P[Z > b] = 0.05$, $P[|z| \le c] = 0.95$를 만족하는 a, b, c를 구해보면 $a = 1.37$, $b = 1.645$, $c = 1.96$임을 알 수 있다.

[예 3.6-5] 확률변수 X의 분포가 정규분포 $N(3, 16)$을 따른다면

$$\begin{aligned} P[4 \le X \le 8] &= P\left[\frac{4-3}{4} \le \frac{X-3}{4} \le \frac{8-3}{4}\right] \\ &= \Phi(1.25) - \Phi(0.25) = 0.8944 - 0.5987 = 0.2957 \end{aligned}$$

[정리 3.6-3] 확률변수 X가 정규분포 $N(\mu, \sigma^2)$을 따른다면, 변환된 확률변수 $V = \frac{(X-\mu)^2}{\sigma^2}$은 자유도가 1인 카이제곱분포를 한다.

증명> 확률변수 $Z = \frac{X-\mu}{\sigma}$는 표준정규분포를 하므로 $V = Z^2$라 두면

$$\begin{aligned} G(v) &= P[Z^2 \le v] = P[-\sqrt{v} \le Z \le \sqrt{v}] \\ &= \int_{-\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= 2\int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \end{aligned}$$

⟨ $z = \sqrt{y}$, $dz = \frac{1}{2\sqrt{y}} dy$로 치환하면⟩

$$= \int_0^{v} \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} dy, \quad 0 \le v$$

이다. 이는 자유도가 1인 카이제곱분포의 분포함수임을 알 수 있다.

[참고] 예이츠의 수정 (Yates's correction; 연속성 수정)

확률변수 X의 분포가 이항분포 $B(n, p)$일 때, 다음의 수정식을 사용하면 확률 값이 비슷해진다.

$$P[a \le X \le b] \approx \Phi\left(\frac{b+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-0.5-np}{\sqrt{np(1-p)}}\right).$$

[정리 3.6-4] 확률변수 $X_1, X_2, \cdots, X_k$가 서로 독립이고 X_i의 분포가 정규분포 $N(\mu_i, \sigma_i^2)$일 때, 확률변수 $Y = a_0 + a_1X_1 + \cdots + a_kX_k$는 평균이 $\mu_Y = a_0 + a_1\mu_1 + \cdots + a_k\mu_k$이고 분산이 $\sigma_Y = a_1^2\sigma_1^2 + \cdots + a_k^2\sigma_k^2$인 정규분포를 갖는다.

증명> 확률변수 Y의 적률생성함수로부터 아래와 같이 증명된다.

$$M_Y(t) = e^{a_0 t} M_{X_1}(a_1 t) \cdot \cdots \cdot M_{X_k}(a_k t)$$

$$\langle M_{X_i}(a_i t) = \exp[\mu_i a_i t + a_i^2\sigma_i^2 t^2/2]\text{이므로}\rangle$$

$$= \exp[(a_0 + a_1\mu_1 + \cdots + a_k\mu_k)t + (a_1^2\sigma_1^2 + \cdots + a_k^2\sigma_k^2)t^2/2]$$

$$= \exp[\mu_Y t + \sigma_Y^2 t^2/2].$$

즉, 확률변수 Y의 적률생성함수는 평균이 μ_Y이고 분산이 σ_Y^2인 정규분포의 적률생성함수이다.

■ 혼합분포

연속형 분포와 이산형 분포의 결합된 형태로 나타나는 분포가 혼합분포(Mixed distribution)이다.

[예] 확률변수 X의 분포함수가 다음과 같이 주어져 있다. 이 분포함수로부터 여러가지 확률과 평균 및 분산을 구해보자.

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{x^2}{4}, & 0 \le x < 1, \\ \frac{1}{2}, & 1 \le x < 2, \\ \frac{x}{3}, & 2 \le x < 3, \\ 1, & 3 \le x. \end{cases}$$

풀이>

$$P[0 < X < 1] = \frac{1}{4}, \quad P[0 < X \le 1] = \frac{1}{2}, \quad P[X = 1] = \frac{1}{4}.$$

$$P[1 \le X \le 2] = \frac{2}{3} - \frac{1}{4} = \frac{5}{12}.$$

$$F'(x) = \begin{cases} \frac{x}{2}, & 0 < x < 1, \\ \frac{1}{3}, & 2 < x < 3. \end{cases}$$

$$P[X = 2] = \frac{1}{6}.$$

$$\mu = E(X) = \int_0^1 x \cdot \frac{x}{2}\,dx + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{6} + \int_2^3 x \cdot \frac{1}{3}\,dx = \frac{19}{12}.$$

$$\sigma^2 = \int_0^1 x^2 \cdot \frac{x}{2}\,dx + 1^2 \cdot \frac{1}{4} + 2^2 \cdot \frac{1}{6} + \int_2^3 x^2 \cdot \frac{1}{3}\,dx - \left(\frac{19}{12}\right)^2 = \frac{31}{48}.$$