연속형 확률 변수와 분포

균일 분포 (Uniform Distribution)

일정한 구간에서 임의로 선택된 수에 대한 확률분포로 이용

균등 분포$X \sim U(a, b)$

$$f(x) = \frac{1}{b-a} I(a < x < b)$$

- $E(X) = \dfrac{a+b}{2} \quad Var(X) = \dfrac{(b-a)^2}{12}$

베타 분포(Beta Distribution)

유한 구간에서 값을 가지는 변수의 분포로 이용

예) 비율의 분포

베타 분포 : $X \sim Beta(\alpha, \beta)$

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1),$$

여기에서

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- $Beta(1, 1) = Uniform(0, 1)$
- $E(X) = \dfrac{\alpha}{\alpha+\beta} \quad Var(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

지수 분포 (Exponential Distribution)

포아송분포를 따르며 발생하는 사건에 대하여 첫번째 사건이 발생할 때까지 대기 시간 (waiting time)의 분포

지수 분포: $X \sim Exp(1/\lambda)$

$$f(x) = \lambda e^{-\lambda x} I(x > 0)$$

$$P(X \leq t) = 1 - P(X > t) = 1 - P(\text{첫 번째 사건이 발생하는 시간} > t)$$
$$= 1 - P([0, t] \text{ 사이에 발생하는 사건의 수} = 0)$$

- 비기억성 (memoryless property): $P(X > t + s | X > s) = P(X > t)$
- $E(X) = 1/\lambda, \ \ Var(X) = 1/\lambda^2$

감마 분포 (Gamma Distribution)

감마 분포: $X \sim Gamma(\alpha, \beta)$

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} I(x > 0)$$

Gamma 함수

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

- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha),\ \Gamma(n) = (n-1)!, \Gamma(1/2) = \sqrt{\pi}$

- $Gamma(1, \beta) = Exp(\beta)$
- $M(t) = \left(\frac{1}{1-\beta t}\right)^\alpha$
- $E(X) = \alpha\beta,\ Var(X) = \alpha\beta^2$

감마 분포 (Gamma Distribution)

- 평균이 λ인 포아송분포를 따르며 발생하는 사건에 대하여 n번째 사건이 발생할 때까지 대기 시간 $\sim Gamma(n, 1/\lambda)$

$$P(X \leq t) = 1 - P(X > t) = 1 - P(n \text{ 번째 사건이 발생하는 시간} > t)$$
$$= 1 - P([0, t] \text{ 사이에 발생하는 사건의 수} \leq n - 1)$$

- 자유도가 n인 카이제곱분포: $\chi_n^2 \sim Gamma\left(\frac{n}{2}, 2\right)$

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

정규분포 (Normal Distribution)

정규분포:, $X \sim N(\mu, \sigma^2)$

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

- $M(t) = e^{\mu t + \sigma^2 t^2 / 2}$
- $E(X) = \mu, \ Var(X) = \sigma^2$
- 표준정규분포 $N(0, 1)$의 누적분포함수 $\Phi(\cdot)$로 표기

이변량 정규분포 (Bivariate Normal Distribution)

이변량 정규분포:, $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$\exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right)\right]$$

- (X, Y) 이변량 정규분포를 따르는 경우, $\rho = 0 \Leftrightarrow X, Y$ 독립

정리 2.1
$f_X(x)$는 연속형 확률변수 X의 확률밀도함수이며 함수 $g(\cdot)$의 역함수가 존재하고 미분 가능할 때 $Y = g(X)$의 확률밀도함수는 다음과 같다.

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

정리 2.2
연속형 확률변수 X의 누적확률분포함수가 $F(\cdot)$일 때 $Y = F(X)$라고 하면, Y는 균일 분포를 따른다. 즉 다음이 성립한다.

$$Y = F(X) \sim Uniform(0, 1)$$

정리 2.3
$U \sim Uniform(0,1)$을 따르는 확률변수이고 $F(\cdot)$은 연속형 확률분포의 누적확률분포함수일 때 $Y = F^{-1}(U)$라고 하면, Y는 누적분포함수 $F(\cdot)$을 갖는 확률변수가 된다.

지수분포 $Exp(1/\lambda)$를 따르는 random number 생성

- 누적분포함수 $F(x) = 1 - e^{-\lambda x}$, 역함수 $F^{-1}(x) = -\frac{1}{\lambda}\log(1-x)$
- $U \sim Uniform(0,1)$에 대하여 $Y = -\frac{1}{\lambda}\log(1-U) \sim Exp(1/\lambda)$

$$-\frac{1}{\lambda}\log(1-U_1), -\frac{1}{\lambda}\log(1-U_2), \ldots, -\frac{1}{\lambda}\log(1-U_n)$$

이변량 확률변수의 변환

$(X, Y) \sim f(x, y)$: 연속형 확률변수

$$U = g_1(X, Y), V = g_2(X, Y)$$

(U, V)의 결합분포?

정리 3.1
$u = g_1(x, y), v = g_2(x, y)$에 대하여 $x = h_1(u, v), y = h_2(u, v)$인 함수 $h_1(\cdot, \cdot)$와 $h_2(\cdot, \cdot)$가 존재하고 이 함수가 미분가능할 때 $U = g_1(X, Y), V = g_2(X, Y)$의 결합확률밀도함수

$$f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J(u, v)|$$

단 $J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial u} h_1(u, v) & \frac{\partial}{\partial v} h_1(u, v) \\ \frac{\partial}{\partial u} h_2(u, v) & \frac{\partial}{\partial v} h_2(u, v) \end{vmatrix}$

NOTE : $X \sim f_X(x)$ 연속형 확률변수일 때 미분가능인 역함수가 존재하는 함수 g에 대하여 $Y = g(X)$의 확률밀도함수

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Example
$X \sim N(0,1),\ \ Y \sim N(0,1)$이고 서로 독립일 때 $U = X + Y, V = X - Y$의 결합확률밀도함수?

- 0부터 1사이의 실수 중에서 임의로 20개를 선택하였을 때, 이 중 0.7보다 큰 숫자가 나올 횟수에 대한 확률분포는?
- 사과 20개가 들어있는 한 상자에서 5개를 임의로 골라 상태를 보고 불량 사과가 없으면 구입할 것이다. 만약 검사할 상자에 4개의 불량 사과가 들어있다고 하자. 그 상자를 구입하지 않을 확률은?
- 기계 A가 1시간 동안 오작동할 확률은 0.02이다. 이 기계가 두 시간 동안 오작동하지 않을 확률은?
- 헌혈 지원자의 80%는 헌혈 가능자라고 한다. 5명의 헌혈지원자 중 적어도 한 명이 헌혈 가능자일 확률은?

Homework Week 5

- 지질학 연구 결과 석유 탐사 중 석유 발견 확률은 0.2이라고 한다. 석유가 3 번 발견될 때까지 탐사를 계속하기로 하였다. 몇 번은 탐사해야할까?
- 만약 낙하산이 A지점과 B지점 임의의 지점에 떨어진다고 하자. 낙하산이 B 보다 A지점에 더 가까이 떨어질 확률은?
- 어느 사람에게 걸려오는 전화 통화 수가 평균적으로 1시간에 0.5통인 포아송분포를 따른다고 할 때 다음 통화가 걸려올 때까지의 대기 시간이 1 시간 이상일 확률은?
- 어느 사람에게 걸려오는 전화 통화 수가 평균적으로 1시간에 0.5통인 포아송분포를 따른다고 할 때 전화 3통이 걸려올 때까지의 대기 시간이 1 시간 이상일 확률은?