05 강

통계적 추론

연속형 확률분포

한국방송통신대학교 통계·데이터과학과
이긍희 교수

학습내용

1. 연속형 균등분포를 이해한다.
2. 지수분포, 감마분포와 베타분포를 이해한다.
3. 정규분포를 이해한다.
4. 로그정규분포와 이변량정규분포를 이해한다.

01

연속형 확률분포의 개요

1 연속형 확률분포의 개요

연속형 확률변수와 확률분포

- **연속형 확률변수**
- **확률밀도함수**

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt,\ \ f(x) = \frac{d}{dx}F(x)$$

- **연속형 균등분포, 정규분포, 로그정규분포, 지수분포, 감마분포, 베타분포**

1 연속형 확률분포의 개요

예 다음 확률밀도함수를 가지는 확률변수 X에서 $P(0.1 \leq X \leq 0.5)$ 과 기댓값을 구하시오.

$$f(x) = 3x^2, 0 < x < 1$$

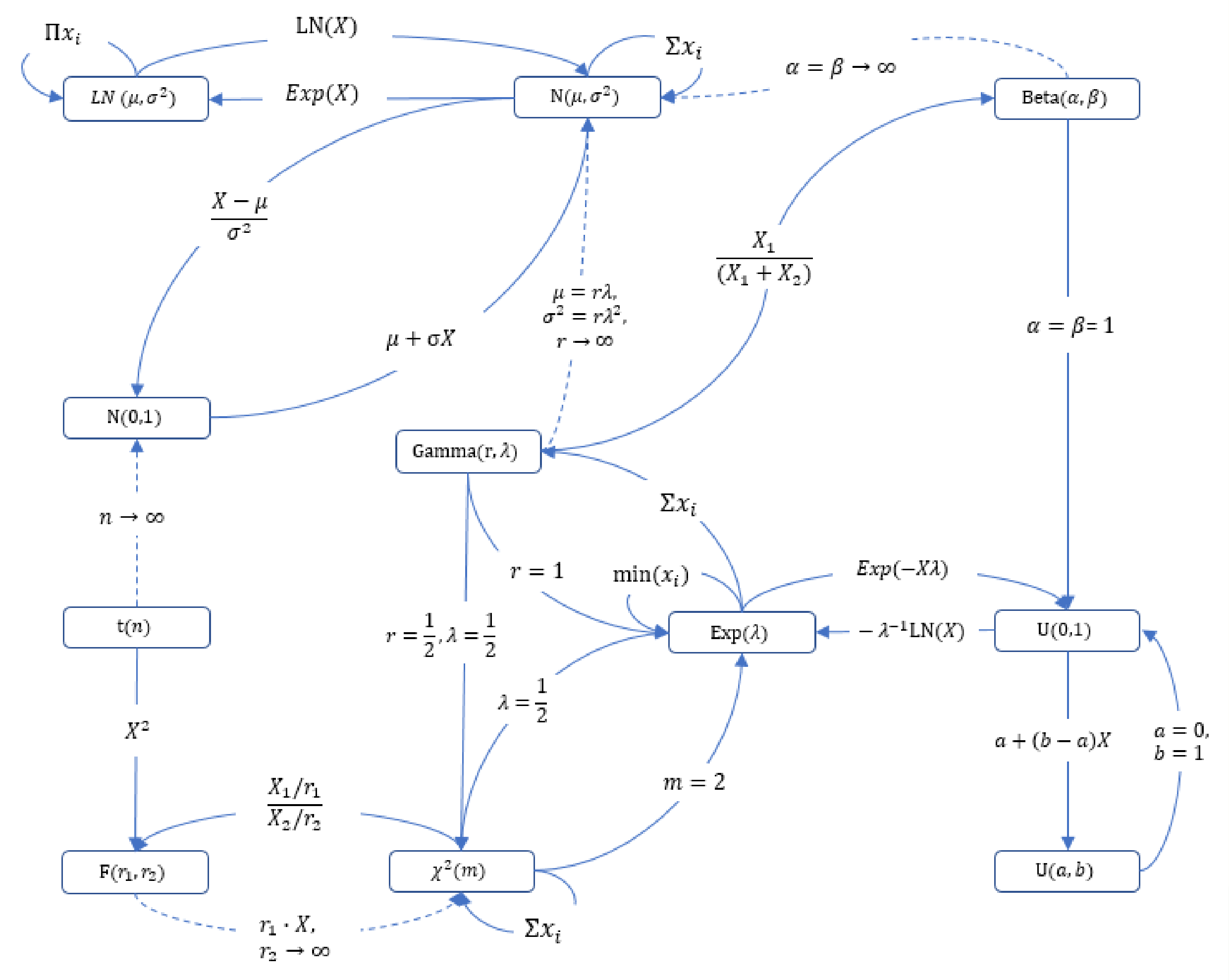
1 연속형 확률분포의 개요

예 다음 확률밀도함수를 가지는 확률변수 X에서 $P(0.1 \leq X \leq 0.5)$과 기댓값을 구하시오.

$$f(x) = 3x^2, 0 < x < 1$$

2 연속형 확률분포의 관계

연속형 확률분포의 관계

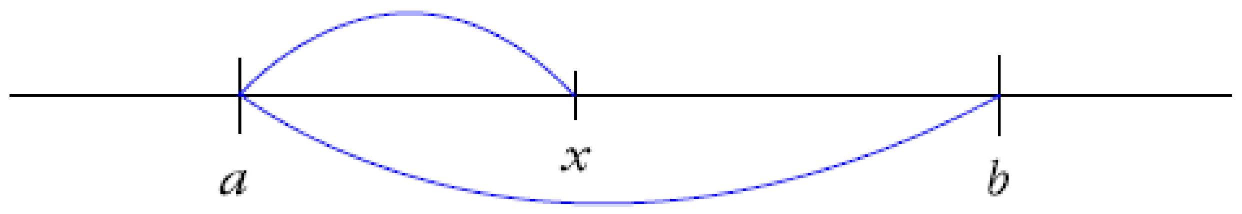


02

연속형 확률분포

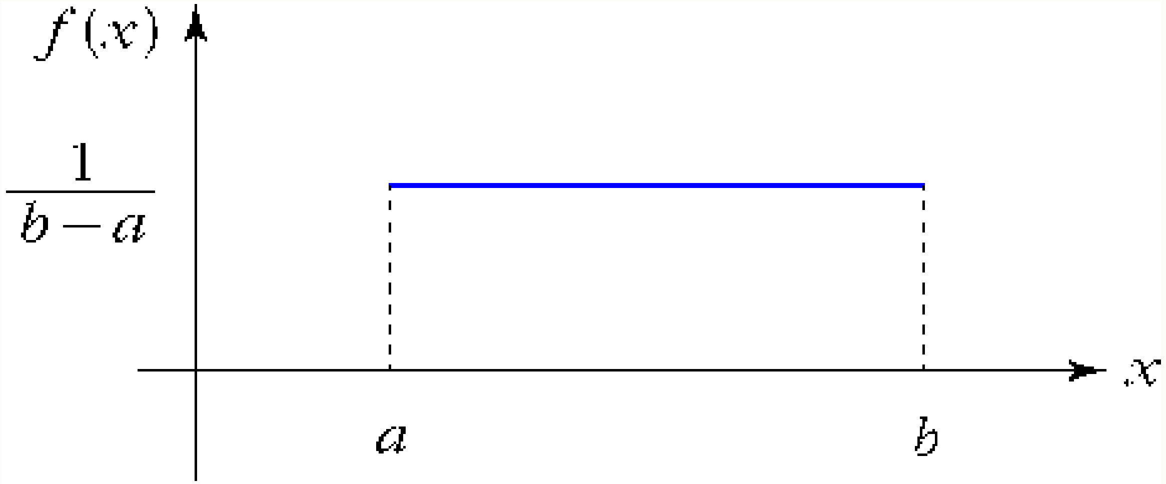
1 연속형 균등분포

- 연속형 균등분포 : $X \sim U(a,b)$



1 연속형 균등분포

- 확률밀도함수 : $U(a,b)$
  - $f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & a > x \text{ or } x > b \end{cases}$



1 연속형 균등분포

- 기댓값 : $U(a,b)$
  - $E(X) = \frac{a+b}{2}$

1 연속형 균등분포

- 분산 : $U(a,b)$
  - $Var(X) = \frac{(b-a)^2}{12}$

1 연속형 균등분포

- 적률생성함수 : $U(a,b)$
  - $M(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$

1 연속형 균등분포

예 $U(0,5)$를 따르는 확률변수 X라고 할 때 $P(2 < X < 4)$는?

2 지수분포

- 지수분포 : $X \sim Exp(\lambda)$
  - **첫 번째 사건이 발생할 때까지 대기시간의 분포**
  - **포아송분포와의 관계**

$$Y \sim Poisson(\lambda t) \quad \Rightarrow \quad P(X \le t) = P(Y \ge 1)$$

2 지수분포

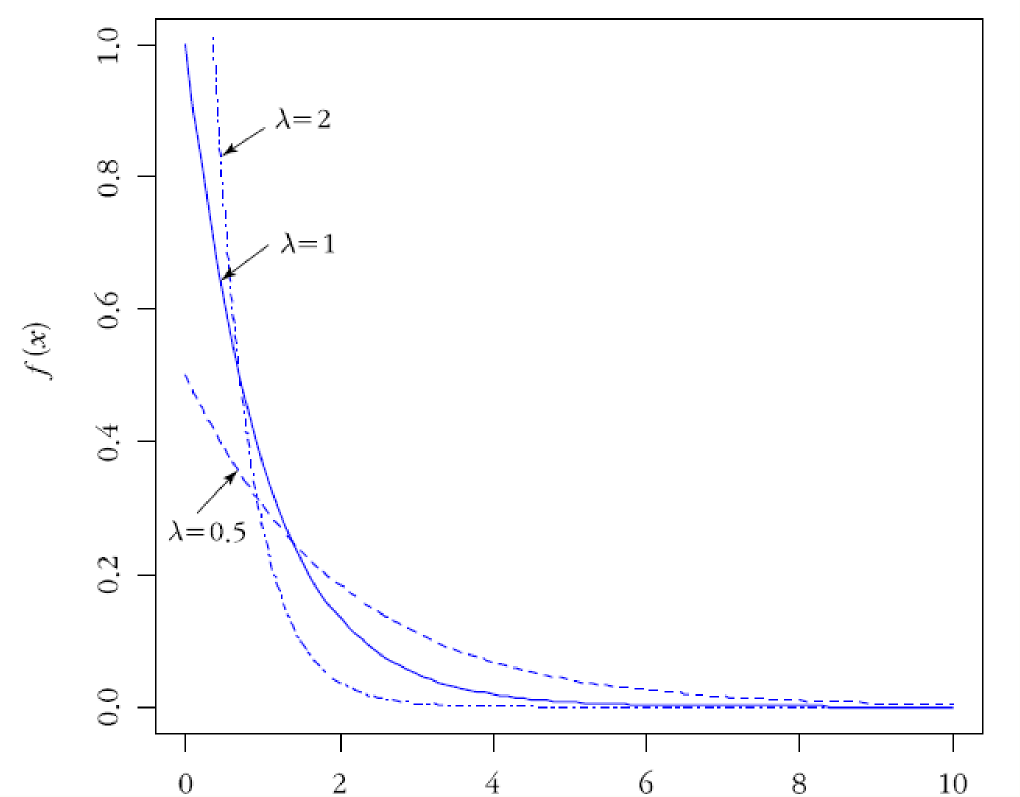
- 누적분포함수 : $Exp(\lambda)$

$$F(t) = P(X \leq t) = \int_0^t \lambda e^{-\lambda x} dx$$

2 지수분포

- 확률밀도함수 : $Exp(\lambda)$

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0$$



2 지수분포

- 기댓값 : $Exp(\lambda)$
  - $E(X) = \frac{1}{\lambda}$
  - $Var(X) = \frac{1}{\lambda^2}$

2 지수분포

- 적률생성함수 : $Exp(\lambda)$

  - $M(t) = \frac{\lambda}{\lambda - t}, \ t < \lambda$

2 지수분포

지수분포의 망각성

- $P(X \geq a + b | X \geq a) = P(X \geq b)$

2 지수분포

예 자동차 수명 $X \sim Exp(1/10)$, 5년 동안 사용해 온 자동차를 앞으로 5년 더 사용할 확률은?

3 베타분포

- 감마분포 : $X \sim Gamma(r, \lambda)$
  - r **번째 현상이 발생할 때까지 대기시간의 분포**
  - $Exp(\lambda) = Gamma(1, \lambda)$
  - **포아송분포와 의 관계**

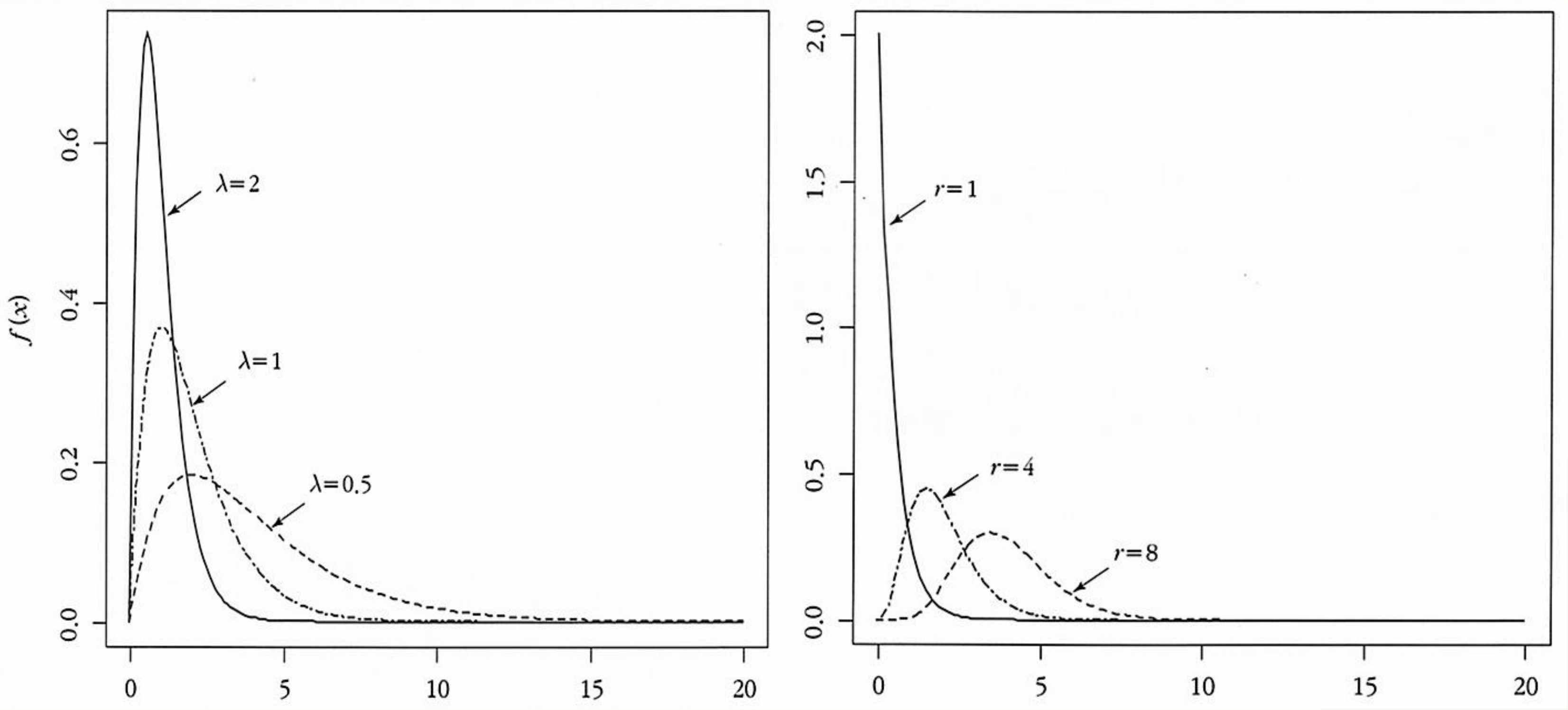
$$Y \sim Poisson(\lambda t) \quad \Rightarrow \quad P(X > t) = P(Y \le r - 1)$$

3 감마분포

- 확률밀도함수 : $Gamma(r, \lambda)$
  - $f(x) = \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}, \quad x \geq 0, \lambda > 0$
  - **감마함수** : $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}\, dy$

3 감마분포

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



3 감마분포

- 적률생성함수 : $Gamma(r, \lambda)$
  - $M(t) = \left(\frac{\lambda}{\lambda - t}\right)^r, t < \lambda$

3 감마분포

- 적률생성함수 : $Gamma(r, \lambda)$
  - $M(t) = \left(\frac{\lambda}{\lambda - t}\right)^r, t < \lambda$

3 감마분포

- 기댓값 : $Gamma(r, \lambda)$
  - $E(X) = \frac{r}{\lambda}$
  - $Var(X) = \frac{r}{\lambda^2}$

3 감마분포

감마분포의 성질

- $Exp(\lambda) = Gamma(1, \lambda)$
- $\chi^2(df) = Gamma\left(\frac{df}{2}, \frac{1}{2}\right)$
- $X_1 \sim Gamma(r_1, \lambda), X_2 \sim Gamma(r_2, \lambda)$

  $\Rightarrow X = X_1 + X_2 \sim Gamma(r_1 + r_2, \lambda)$

3 감마분포

예 1년 평균 고장률이 0.1인 기기의 두 번째 고장이 2년 이후 일어날 확률은?

4 베타분포

- 베타분포 : $X \sim Beta(\alpha, \beta)$
  - **유한구간의 확률현상을 모형화하기 위한 분포로 베이즈 추론에서 자주 이용하는 분포**
  - $X_1 \sim Gamma(\alpha, \lambda), X_2 \sim Gamma(\beta, \lambda)$

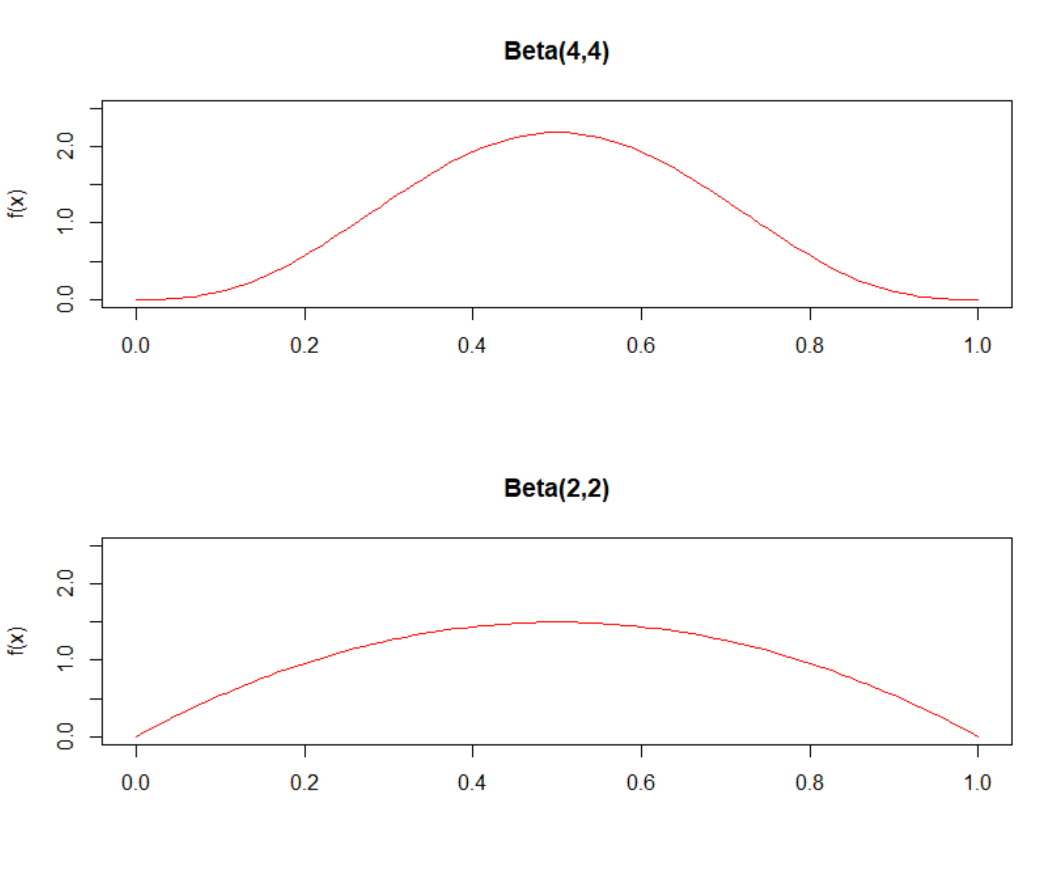
$$\Rightarrow X = \frac{X_1}{X_1 + X_2} \sim Beta(\alpha, \beta)$$

4 베타분포

- 확률밀도함수 : $Beta(\alpha, \beta)$
  - $f(x) = \frac{1}{Beta(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1$
  - $Beta(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$

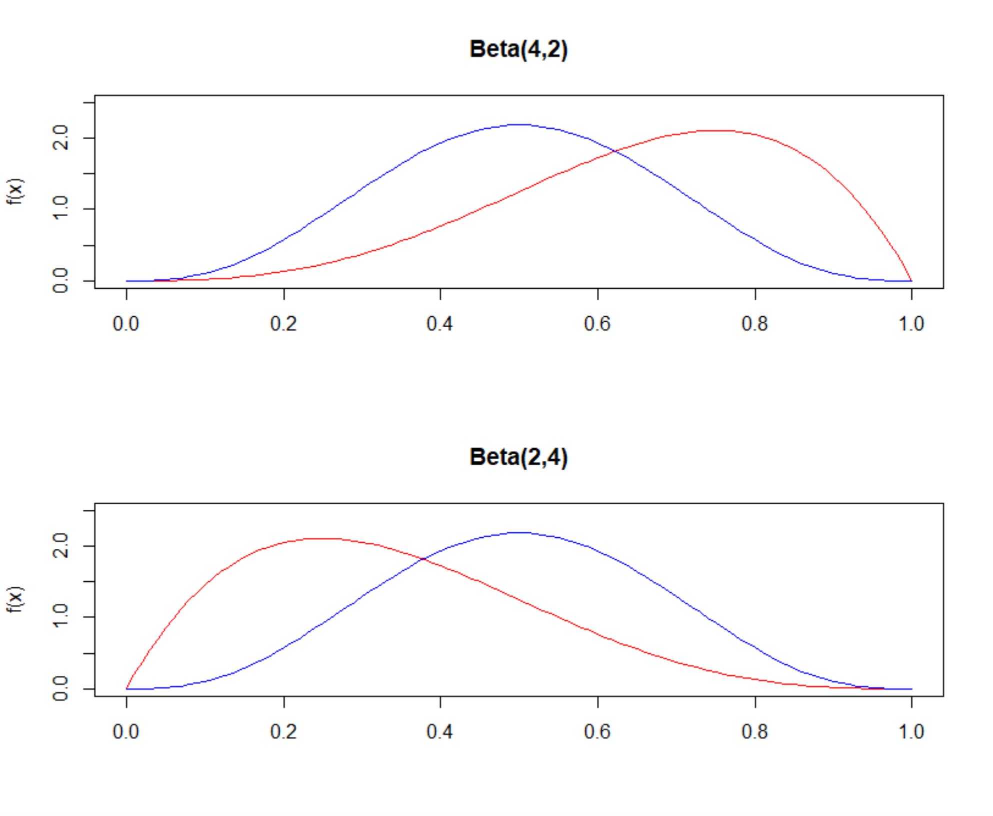
4 베타분포

- 확률밀도함수 : $Beta(\alpha, \beta)$
  - $f(x) = \frac{1}{Beta(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1$



4 베타분포

- 확률밀도함수 : $Beta(\alpha, \beta)$
  - $f(x) = \dfrac{1}{Beta(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1$

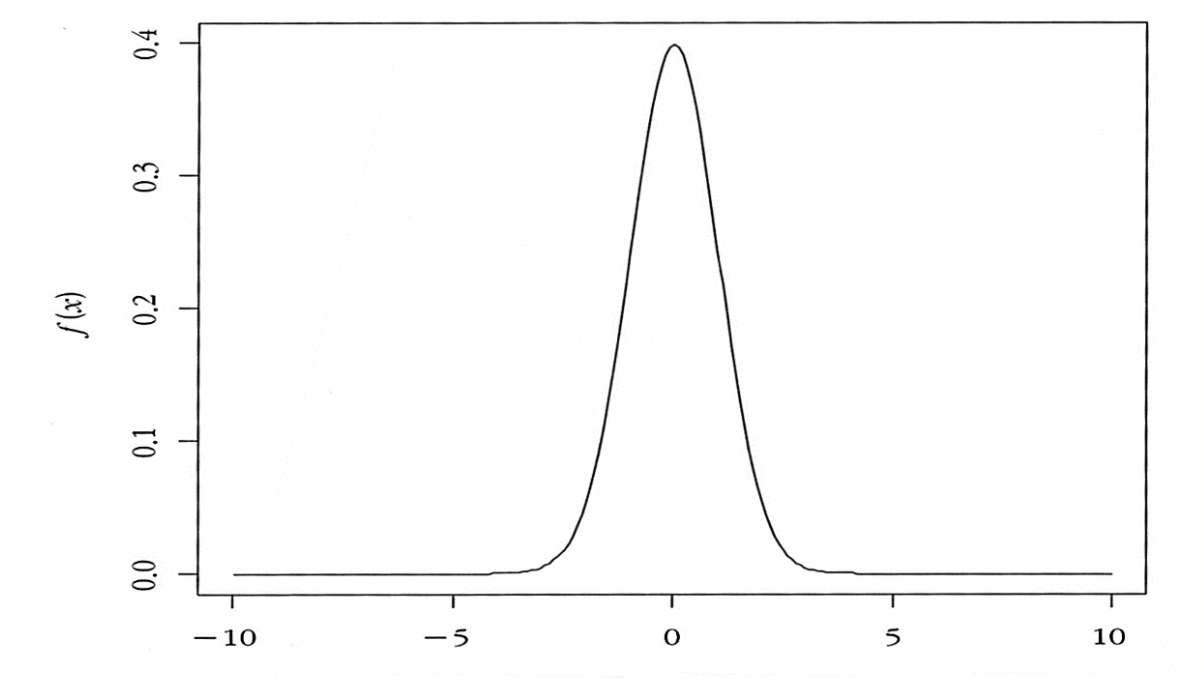


4 베타분포

- 기댓값 : $Beta(\alpha, \beta)$
  - $E(X) = \frac{\alpha}{\alpha+\beta}$
  - $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

5 정규분포

- 정규분포 : $X \sim N(\mu, \sigma^2)$
  - 종모양의 좌우 대칭인 분포
  - 평균과 분산으로 그 형태 결정



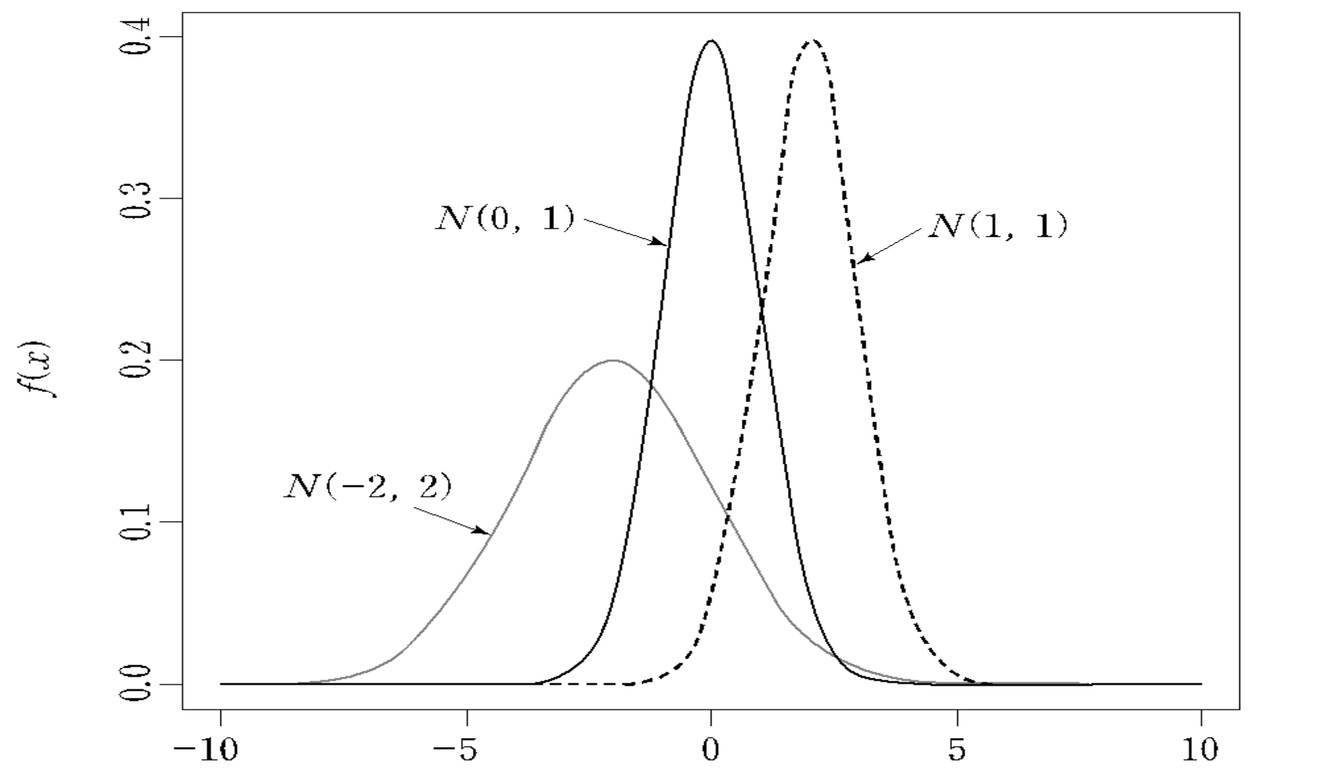
5 정규분포

역 사

- **1738년 : 드 무아브르**(Abraham de Moivre)는 **이항분포 극한을 정규분포의 수학적 형태로 표현**
- **1810년 : 라플라스**(Pierre-Simon Laplace) 중심극한정리
- **1823년 : 가우스**(Carl Friedrich Gauss)는 **천문 관측 데이터를 정규분포로 설명(오차이론)**

5 정규분포

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



5 정규분포

표준정규분포

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

5 정규분포

- 정규분포 $N(\mu, \sigma^2)$의 표준화
  - $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

5 정규분포

- 기댓값 : $N(\mu, \sigma^2)$
  - $E(X) = \mu$
  - $Var(X) = \sigma^2$

5 정규분포

- 적률생성함수 : $N(\mu, \sigma^2)$
  - $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

5 정규분포

- 적률생성함수 : $N(\mu, \sigma^2)$
  - $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

5 정규분포

표준정규분포의 누적분포함수

- $\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] du$

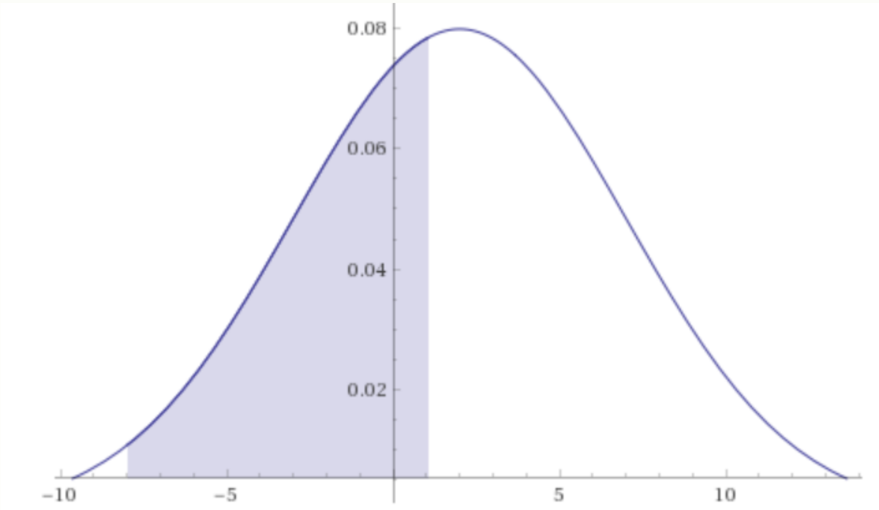
z	0	1	1.645	1.96	2	3
P(Z<z)	0.5000	0.8413	0.9500	0.9750	0.9772	0.9887

5 정규분포

- 정규분포 $N(\mu, \sigma^2)$의 계산
  - $X \sim N(\mu, \sigma^2)$
  - $P(c \le X \le d) = \Phi\left(\frac{d-\mu}{\sigma}\right) - \Phi\left(\frac{c-\mu}{\sigma}\right)$

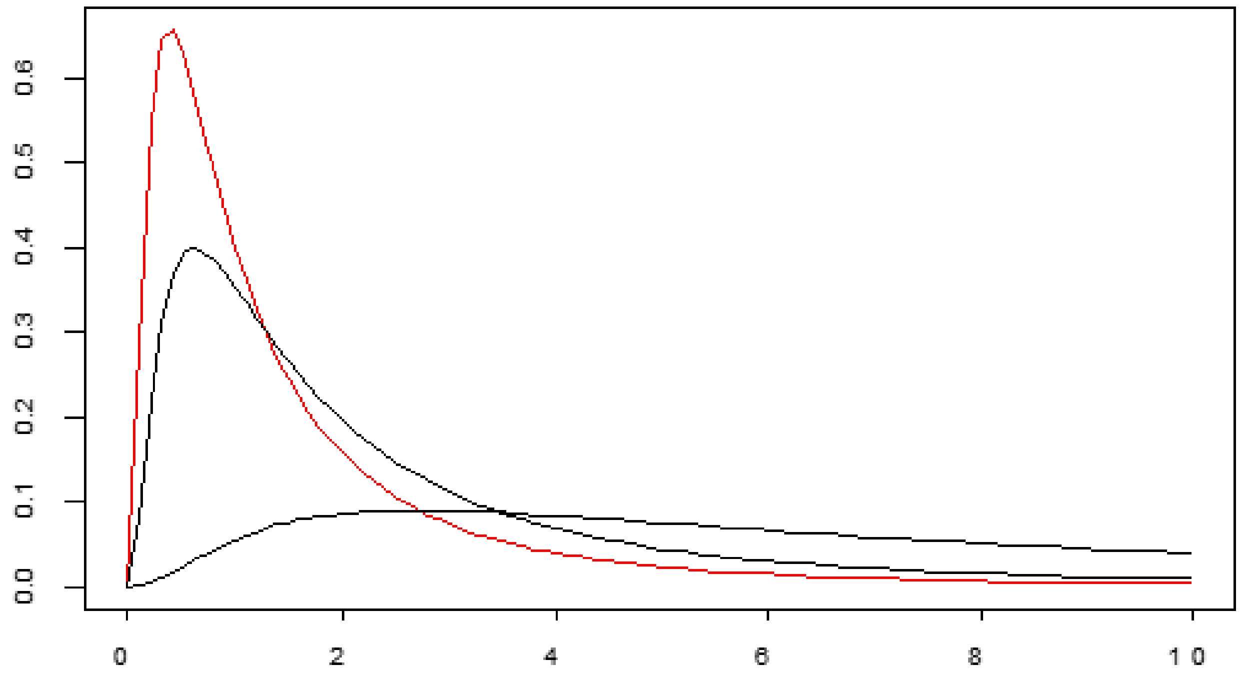
5 정규분포

예 $X \sim N(2,25)$일 때 $P(-8 < X < 1)$ 은?



6 로그정규분포

- 로그정규분포 : $X \sim Lognormal(\mu, \sigma^2)$
  - $\mathrm{Y} = \log(X)$, $Y \sim N(\mu, \sigma^2)$
  - $f(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left[-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right]$, $x > 0$



6 로그정규분포

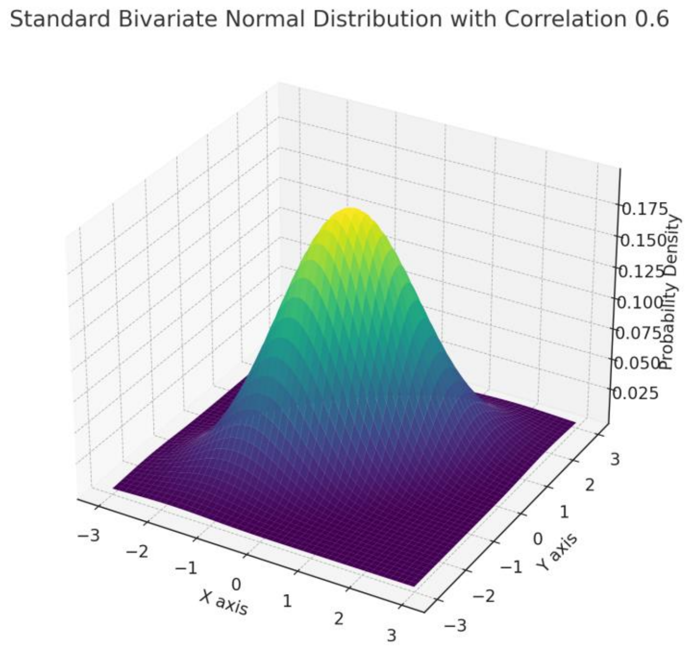
- 로그정규분포 : $X \sim Lognormal(\mu, \sigma^2)$
  - $\mathrm{Y} = \log(X)\,, Y \sim N(\mu, \sigma^2)$
  - $f(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left[-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right], x > 0$

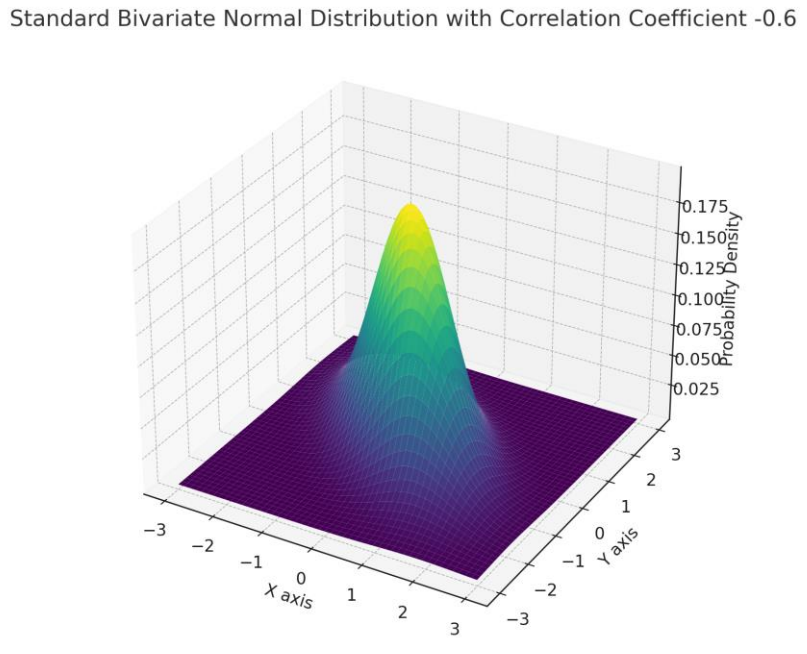
6 로그정규분포

- 기댓값 : $Lognormal(\mu, \sigma^2)$
  - $E(X) = e^{\mu+\sigma^2/2}$
  - $Var(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$

7 이변량 정규분포

- 표준 이변량 정규분포 : $(Z_1, Z_2) \sim N_2(0,0,1,1,\rho)$
  - $f(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[z_1^2 - 2\rho z_1 z_2 + z_2^2\right]\right]$





7 이변량 정규분포

- 이변량 정규분포 : $(X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$
  - $f(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right]$

7 이변량 정규분포

- 적률생성함수 : $(X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$
  - $M(t_1, t_2) = \exp\left[\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}[\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2]\right]$

7 이변량 정규분포

이변량 정규분포의 성질

- $(X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

- $X_1 \sim N(\mu_1, \sigma_1^2),\ X_2 \sim N(\mu_2, \sigma_2^2)$

- $X_2 | X_1 = x \sim N\left(\mu_2 + \frac{\rho\sigma_2(x-\mu_1)}{\sigma_1}, (1-\rho^2)\sigma_2^2\right)$

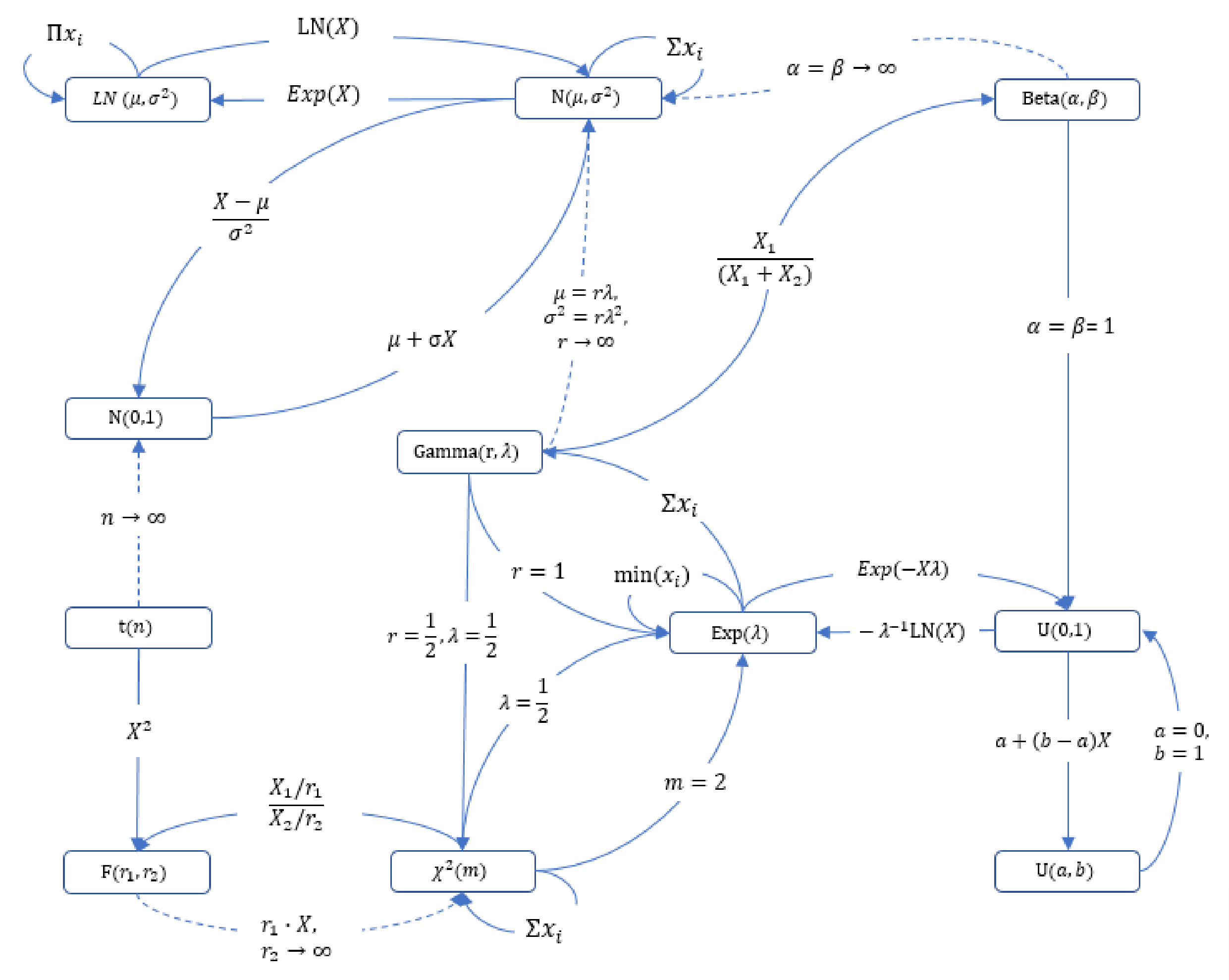
7 이변량 정규분포

이변량 정규분포의 성질

- $(X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$
- $Corr(X_1, X_2) = \rho$
- $\rho = 0 \Leftrightarrow X_1, X_2$ **독립**

8 연속형 확률분포의 관계

연속형 확률분포의 관계



정리하기

- 연속형 균등분포는 확률변수가 구간 각 값을 가질 가능성이 같을 때 분포이다.
- 지수분포는 사건 발생 때까지 기다리는 시간의 분포이다.
- 감마분포는 r 번째 현상이 발생할 때까지 대기시간의 분포이다.
- 베타분포는 유한구간의 확률현상을 모형화하기 위한 분포로 베이즈 추론에서 자주 이용하는 분포이다.

정리하기

- 정규분포는 종 모양의 좌우 대칭인 곡선형태의 분포로 평균과 분산으로 그 형태를 결정할 수 있다.
- 로그정규분포는 로그변환된 확률변수가 정규분포를 따를 때의 분포이다.
- 이변량 정규분포는 관련 있는 두 확률변수가 각각 정규분포를 따를 때 두 확률변수의 결합확률분포이다.

05 강

다음시간 안내

표본분포 1