혼합분포

정의

- 한 확률변수의 분포가 다른 분포에 영향을 받아 변형될 때 이를 혼합분포라고 한다.

1. pdf가 $f_1(x), f_2(x), \cdots, f_n(x)$이고, 범위는 $s_1, s_2, \cdots, s_n$ 이고,
평균은 $\mu_1, \mu_2, \cdots, \mu_n$ 이고 분산은 $\sigma_1, \sigma_2, \cdots, \sigma_n$ 일 때

1) 이 확률변수의

(1) 혼합된 $PDF\ f(x) = p_1 f_1(x) + p_2 f_2(x) + \cdots + p_n f_n(x)$이다. 즉, 확률로 가중된 가중합

(2) 혼합된 평균은 $\mu = p_1\mu_1 + p_2\mu_2 + \cdots + p_n\mu_n$ 즉, 확률로 가중된 가중합

(3) 혼합된 분산은

- var(x) = $p_1{\sigma_1}^2 + p_2{\sigma_2}^2 + \cdots + p_n{\sigma_n}^2$ 에서
- $\sum_{i=1}^{n} p_i \int (x-\bar{\mu})^2 f_i(x)\,dx$ 이므로, 변환하면
- $\sum_{i=1}^{n} p_i \int [(x-\mu)+(\mu-\bar{\mu})]^2 f_i(x)\,dx$
$= \sum_{i=1}^{n} p_i \int [(x-\mu)^2 + 2(x-\mu)(\mu-\bar{\mu}) + (\mu-\bar{\mu})^2] f_i(x)\,dx$

정의

- 한 확률변수의 분포가 다른 분포에 영향을 받아 변형될 때 이를 혼합분포라고 한다.
- $\sum_{i=1}^{n} p_i \left[\int (x-\mu)^2 f_i(x)dx + \int 2(x-\mu)(\mu-\bar{\mu}) f_i(x)dx + \int (\mu-\bar{\mu})^2 f_i(x)dx\right]$
- 이 때, $\int 2(x-\mu)(\mu-\bar{\mu}) f_i(x)dx$ = E[$(x-\mu)$] = 0 이므로
- $\sum_{i=1}^{n} p_i \left[\int (x-\mu)^2 f_i(x)dx + \int (\mu-\bar{\mu})^2 f_i(x)dx\right]$
  $= \sum_{i=1}^{n} p_i \sigma^2 + \sum_{i=1}^{n} p_i (\mu-\bar{\mu})^2$
- 즉, 분산의 가중합과 총분산(총평균 – 평균벡터) 의 가중합의 합이다.

정의

- 무한개의 분포의 합으로 확정

1. 적분을 이용하여 pdf를 표현할 수 있다.

1) 이 때, 어떤 분포의 모수가 다른 확률분포에 의존하는 분포의 통계량을 구할 수 있다.
2) 예를 들어, 푸아송 분포의 평균 θ 가 감마분포를 따르는 복합분포의 경우,

(1) $p(x|\theta)g(\theta) = \frac{(\theta)^x \boldsymbol{e}^{-\theta}}{\boldsymbol{x}!} \cdot \frac{1}{\Gamma(a)\beta^a} \theta^{a-1} e^{-\frac{\theta}{\beta}}$

(2) 이 식을 θ에 대해 적분하면 x의 주변 pdf를 구할 수 있다.

- $p(\boldsymbol{x}) = \int_{\mathbf{0}}^{\infty} \frac{(\theta)^x \boldsymbol{e}^{-\theta}}{\boldsymbol{x}!} \cdot \frac{1}{\Gamma(a)\beta^a} \theta^{a-1} e^{-\frac{\theta}{\beta}} \boldsymbol{d}\theta$

예제

- X 가 $x = 1,2,\cdots,n$ 일 때 $\theta(\mathbf{1}-\theta)^{\boldsymbol{x}-\mathbf{1}}$인 조건부 기하 pdf를 갖고, θ가 베타pdf를 따를 때 x 의 주변 pdf를 구하라.

1. 복합 pdf를 구하면

1) $f(x|\theta)g(\theta) = \theta(\mathbf{1}-\theta)^{\boldsymbol{x}-\mathbf{1}} \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \theta^{a-1} (1-\theta)^{\beta-1}$

2) 이를 정리하면

(1) $f(x|\theta)g(\theta) = \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \theta^{a} (1-\theta)^{x+\beta-2}$

2. 무조건부 pdf는

1) $\int_{\mathbf{0}}^{\infty} \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \theta^{a} (1-\theta)^{x+\beta-2} \boldsymbol{d}\theta = \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \int_{\mathbf{0}}^{\infty} \theta^{a} (1-\theta)^{x+\beta-2} \boldsymbol{d}\theta = \frac{\Gamma(a+\beta)}{\Gamma(a)\Gamma(\beta)} \cdot \frac{\Gamma(a+1)\Gamma(\beta+x-1)}{\Gamma(a+\beta+x)}$