4장. 가능도 (Likelihood) : 개념 · 계산 · 속성

3장까지는 확률만 다루었습니다만 이 장부터 통계적 추론을 시작합니다. 확률의 장(場)에서 통계의 마당으로 넘어오는 길목에 가능도(可能度, 또는 尤度, likelihood)가 있습니다. 그러므로 가능도의 역할은 통계학에서 굉장히 중요합니다. 겉보기엔 가능도가 곧 확률(밀도)인 것 같지만 철학적 위상은 서로 매우 다릅니다. 확률이 연역적이라면 가능도는 귀납적입니다. 4.1절에서는 가능도 함수를 그려보고 감상하고 탐색해봅니다. 봉우리를 찾아 올라가기도 하고 일정 높이에 금을 그어 잘라보기도 합니다.

이와 같이 가능도를 탐색하는 데 있어 순수한 해석적인 방법으로는 해결할 수 없는 실제 사례들이 많습니다. 그런 때는 컴퓨터의 계산력을 활용하는 수치적 방법을 사용합니다. 4.2절에서는 이분법(bisection method)과 뉴튼-라프슨(Newton-Raphson) 방법을 근(根) 구하기와 최대점 구하기에 쓸 것입니다.

4.3절에서는 가능도의 여러 사례들을 다룹니다. 어떤 것은 특수하면서도 전형적입니다. 또한 파라미터가 2개(이상)인 경우도 다룹니다. 그래프로도 표현해봅니다. 뉴튼-라프슨(Newton-Raphson) 방법을 확장해 적용해봅니다.

충분통계량은 파라미터 θ에 대한 핵심적 정보를 갖는 통계량입니다. 이것만 있으면 원자료가 없어도 됩니다. 그러므로 충분성은 자료축소로 연결됩니다. 4.4절에서는, '통계 게임'의 맥락에서 충분성의 개념을 파헤쳐 설명합니다. 또한 충분통계량을 탐지해 내는 데 유용한 인수화(factorization) 정리를 제시합니다.

4.5절에서는 충분통계량의 여러 사례를 봅니다. 파라미터가 2개(이상)인 경우 이들 파라미터를 위해 결합적으로 충분한 통계량에 대하여 논의합니다. 인수화 정리도 확장해 제시합니다. 또한 지수족(exponential family)의 확률분포에서 충분통계량에 대하여도 설명하기로 하겠습니다.

차례 :

4.1 가능도 (Likelihood)

모집단(母集團, population)에 대해 알기 위해 우리가 하는 전형적인 일은 그 집단에서 임의표본을 취하여 관측하는 것입니다. 여기서는 모집단을

$$f(x;\theta)$$

로 확률모형화 하겠습니다. 여기서 x는 자료값이고 θ는 소위 파라미터(parameter), 또는 모수(母數)라고 하는데 이것은 개별 상황에 적응시킬 수 있는 "자유" 상수입니다. 그리고 임의표본의 자료값들을 $x_1, \cdots, x_n$이라고 합시다.

그러면 표본 자료값들에 대한 결합확률(이산형 변수의 경우), 또는 결합확률밀도(연속형 변수의 경우)는

$$f(x_1;\theta)\cdots f(x_n;\theta), \quad 즉 \quad \prod_{i=1}^{n} f(x_i;\theta) \tag{1}$$

가 됩니다. 모집단에 대한 정보의 근원(根源)은 (1)이 어떤 θ에 대하여는 크고 어떤 θ에 대하여는 작다는 데 있습니다. 왜 그런가요? 만약 (1)이 확률모형이 허용하는 모든 θ에 대하여 동일하다면 표본 데이터 $x_1, \cdots, x_n$이 θ를 변별해내는 데 전혀 도움이 되지 못한다는 의미가 되기 때문입니다.

예를 들어보겠습니다. 어느 보험회사 대리점에서 매월 보험청구 건수를 조사한 결과 $3, 10, 5, 4\,(= x_1, x_2, x_3, x_4)$건이었다고 합시다. 그런데 보험청구건수를 포아송 분포 Poisson(θ)로 모형화한다고 합시다. 즉,

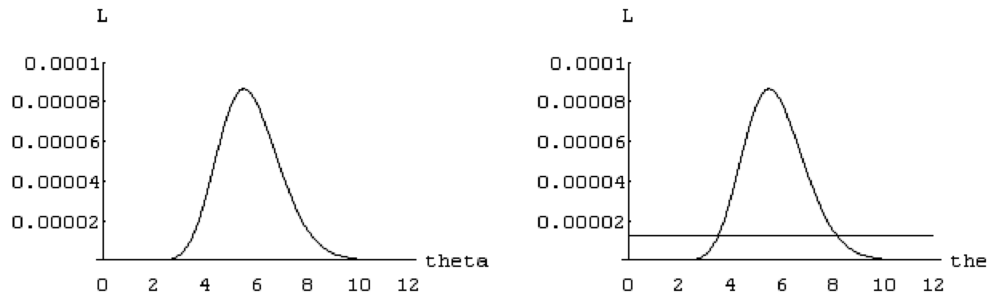
$$f(x;\theta) = \theta^x e^{-\theta} / x!,\ \theta > 0,\ x = 0, 1, 2, 3, \cdots$$

이므로 결합확률 (1)은

$$\prod_{i=1}^{4} f(x_i;\theta) = \frac{\theta^{22} e^{-4\theta}}{3!\,10!\,5!\,4!} \tag{2}$$

가 됩니다. (2)를 θ의 두 값인 $\theta = 4$와 $\theta = 10$에 대하여 계산해보겠습니다. (약간의 품이 들어간 후에…) 그 결과는 각각 3.16×10^{-5}와 6.78×10^{-7}입니다. 그러니 앞의 값이 뒤의 값의 46.6배나 됩니다. 따라서 $\theta = 4$가 $\theta = 10$에 비하여 훨씬 더 그럴듯한(more plausible) 파라미터라는 것을 직관적으로 알 수 있습니다. 그러면 $\theta = 4$보다 더 그럴듯한 파라미터가 있을까요? 그것보다 더, 그것보다 더, … ?

이런 물음에 답하기 위해서는 (2)를 θ에 관한 함수로 놓고 그것을 플롯해보는 것이 좋겠습니다. <그림 1>의 왼쪽 것을 보십시오. (2)가 $\theta = 5$와 6사이의 어떤 값에서 가장 큰 값을 취하고 $\theta = 2$ 이하나 11이상에서는 거의 0에 가까운 값이 되는



<그림 1> 포아송 표본의 가능도

것을 볼 수 있습니다. (2)와 같은 θ의 함수를 가능도 함수(可能度函數, likelihood function)라고 합니다. 앞으로는 가능도 함수를 간략히 '가능도'라고 하겠습니다. 많은 통계학 서적에서는 가능도 대신 우도(尤度)라는 용어를 쓰고 있으므로 여러분도 두 용어를 다 알아둘 필요가 있습니다.

<u>가능도</u>(可能度, likelihood) : 임의표본에 대한 정의는 다음과 같습니다.

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

이렇게 정의되는 가능도는 스스로 합리적 믿음의 정도(degree of rational belief)를 상대적으로 나타낸다고 할 수 있습니다. 따라서 가능도를 그려놓고 모양을 감상하고 특징을 탐색하는 일에서 통계적 추론이 시작됩니다. 사람한테는 눈이 인상을 좌지우지하듯이, 가능도의 대표적인 특성은 봉우리에 있습니다. 따라서 가능도를 가장 크게 하는 파리미터를 찾아 $\hat{\theta}$로 표기합시다. 즉,

$$L(\hat{\theta}) = \max_{\theta} L(\theta)$$

입니다. 이 $\hat{\theta}$을 θ에 대한 최대가능도추정값(maximum likelihood estimate, mle)이라고 합니다. 이것은 한 점이기 때문에 점 추정(point estimation)의 일종입니다.

θ의 mle $\hat{\theta}$을 구하는 방법에는 두 가지가 있습니다. 한 방법은 가능도 $L(\theta)$의 최대값을 해석적으로 찾는 것입니다. 주로, 가능도 $L(\theta)$에 로그(logarithm)를 취한 뒤 미분을 이용하게 됩니다. 예로서, 가능도 (2)를 최대화하여 봅시다.

$$L(\theta) = \frac{\theta^{22} e^{-4\theta}}{3!\,10!\,5!\,4!}$$

이므로 로그를 취하면

$$\log_e L(\theta) = 22\log\theta - 4\theta + \text{constant}$$

입니다. 따라서 양변을 θ로 미분하면

$$\frac{d}{d\theta}\log_e L(\theta)\left(= \frac{L'(\theta)}{L(\theta)}\right) = \frac{22}{\theta} - 4$$

가 됩니다. 그러므로

$$L'(\theta) = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{22}{4} = 5.5$$

입니다. <그림 1>에서 가능도의 최대값은 $\hat{\theta} = 5.5$ 에서 나옵니다.

앞의 풀이 과정에서 나온 표현인 가능도 $L(\theta)$의 로그는 일반적으로 유용한 함수입니다. 그래서 특별히 다음과 같이 정의합니다.

<u>로그 가능도</u>(log-likelihood) $l(\theta)$: 정의

$$l(\theta) = \log_e L(\theta).$$

임의표본에 대하여 로그 가능도는

$$l(\theta) = \sum_{i=1}^{n}\log_e f(x_i;\theta).$$

그러므로 가능도 대신에 로그 가능도를 다루는 것이 수학적으로 동등하면서도 실제적으로 유용한 경우가 많습니다.

참 θ 값은 우리가 표본 데이터로부터는 확실히 알 수 없기는 해도 꼭 $\hat{\theta}$과 일치한다고는 볼 수 없습니다. 그래도 그 근처일 것이라는 추측은 가능하지요. 그러므로 $\hat{\theta}$을 중심으로 가능성이 집중된 구간을 정의할 필요가 있습니다.

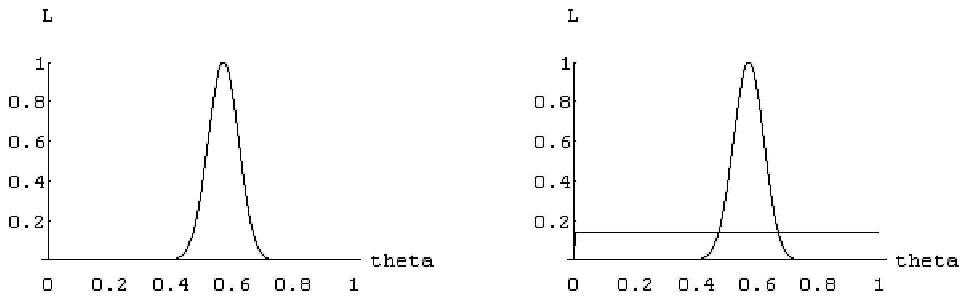
<u>k 분의 1 규칙</u> : 가능성이 집중된 구간을 다음과 같이 정의합시다.

$$C_k = \left\{\theta \mid L(\theta) \geq \frac{1}{k}\max_\theta L(\theta)\right\}, \text{ for } k \geq 1.$$

<그림 1>의 오른 쪽 그림에서 수평선은 $\frac{1}{7}L(\hat{\theta})$에 그어졌습니다 (즉, $k=7$). 수평선과 가능도와의 교점을 찾아보면 $\theta = 3.50, 8.15$ 입니다 (그 구체적 방법에 대하여는 다음 절에서 설명하겠습니다). 따라서

$$C_7 : \theta \in (3.50, 8.15)$$

입니다. k분의 1 규칙을 통하여 이와 같이 일종의 구간 추정(interval estimation)을 할 수 있습니다.



<그림 2> 베르누이 표본의 정규화 가능도

k의 선택은 직관에 의존하여 결정할 수밖에 없습니다만 $k=7$ 정도면 어떨까요? 최대 가능도의 1/7 이상이 되는 가능도 값을 가져야 고려해볼 만한 파라미터라고 보자는 것입니다 (이에 대한 수리적 정당화는 7장에서 하겠습니다).

다른 예를 하나 더 들어보겠습니다. K 대학교에서 특정 학내 이슈에 대해 100명을 임의추출하여 여론 조사를 하였습니다. 그랬더니 57명이 찬성, 43명이 반대하였습니다. 이 때 알고자 하는 것은 K 대학교 전체 학생들(모집단)의 찬성률 θ입니다.

이 경우를 베르누이 분포로 모형화합시다. 즉 $x_1, \cdots, x_n\,(n=100)$을 베르누이 분포 Bernoulli$(\theta)$로부터의 실현된 자료값이라고 합시다. 이 때 가능도는

$$L(\theta) = \theta^{x_1+\cdots+x_n}(1-\theta)^{1-x_1+\cdots+1-x_n} = \theta^s(1-\theta)^{n-s}, \quad 0<\theta<1 \qquad (3)$$

입니다 (여기서, $s = x_1+\cdots+x_n$). 따라서 로그 가능도 $l(\theta)$는

$$l(\theta) = s\log_e\theta + (n-s)\log_e(1-\theta)$$

가 됩니다. 그러므로

$$\frac{d}{d\theta}l(\theta) = \frac{s}{\theta} - \frac{n-s}{1-\theta} = \frac{s-n\theta}{\theta(1-\theta)}$$

이고 $\hat\theta = \frac{s}{n}$ 입니다. 이 자료에서는 $s=57, n=100$이므로 $\hat\theta = 0.57$입니다.

그런데 n이 크면 자연스럽게 $\max_\theta L(\theta)$는 값 자체가 매우 작아집니다. 이 경우엔 2.1×10^{-30}입니다. 그러므로 가능도를 직접 플롯하는 대신 표준화된 형태의 가능도인

정규화 가능도 (정의) : $L(\theta)/\max_\theta L(\theta)$

가 더 좋을 때가 많습니다. <그림 2>는 (3)을 정규화된 형태로 플롯한 것입니다. 그리고 1/7 규칙을 적용하여 θ에 대한 구간을 구하여 본 결과는

$$C_7 \;:\; \theta \in (0.471, 0.665)$$

입니다. 즉, 1/7 규칙에 의한 찬성률 θ에 대한 구간은 $0.471 \le \theta \le 0.665$ 입니다.

4.2 수치적으로 근(根) 구하기 · 최대점 구하기

컴퓨터가 흔해졌고 좋아졌다지만 실제로 그만큼 활용하지 않으면 무슨 소용이 있겠습니까? 수리통계학에서도 마찬가지입니다.

통계적 방법론의 틀이 만들어진 시기는 1920년 정도부터 약 50년간인데 이 때의 계산능력은 지금과는 비교가 안되었습니다. 그랬기 때문에 옛 학자들은 뛰어난 선견지명을 갖고 있었음에도 불구하고 계산을 많이 요구하는 방법보다는 계산이 간편한 방법을 선호하였던 것입니다. 앞 절에서의 k분의 1 규칙이 좋은 한 예입니다. R.A. Fisher(1890-1962)는 그의 1956년 저서인 <u>Statistical Methods and Scientific Inference</u> (Chapter III, Section 6)에서 $1/2, 1/5, 1/15$ 규칙을 사용한 구간추정의 방법을 제안하였지만 그 자신도 실용적으로 사용하지는 않은 듯합니다. 계산이 어려운 것이 주된 이유였을 것입니다.

k분의 1 규칙에 의해 파라미터 θ에 대한 구간을 구하려면

$$L(\theta) = \frac{1}{k} \max_\theta L(\theta),$$

즉

$$l(\theta) - \max_\theta l(\theta) + \log_e k = 0 \quad (k \ge 1)$$

을 풀어야 합니다. 이 절에서는 연속함수 $g(.)$의 근 θ를 구하는 방법을 설명하기로 합니다 (즉, $g(\theta) = 0$).

여러 방법이 있지만 여기서는 그 중에서 이분법은 근이 있을 것으로 생각되는 구간을 절반씩 쪼개어 점점 좁혀 가는 방법입니다. 그 원리는 다음과 같습니다.

<u>이분법</u>(二分法, bisection method)

함수 $g(\theta)$가 연속이라고 합시다. 구간(θ_1, θ_2)의 양끝 점에서 계산한 함수값 $g(\theta_1)$과 $g(\theta_2)$의 부호가 다르면, 즉

$$g(\theta_1)\, g(\theta_2) < 0$$

을 만족하면 이 구간에 근이 적어도 1개 존재합니다. $g(\theta)$가 연속이라고 했기 때문이죠. 다음으로 구간의 중간점 $\theta = (\theta_1 + \theta_2)/2$에서 함수값 $g(\theta)$를

<표 1> 이분법으로 $1/k$ 규칙에 의한 구간 구하기

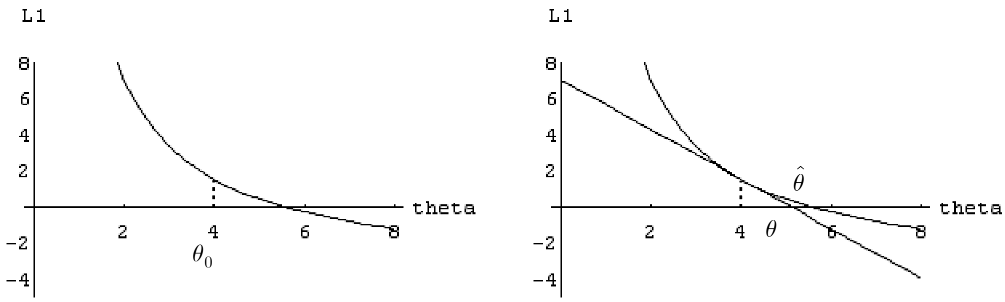
```
/* Bisection method for 1/k rule interval computing */
/* bike.iml                                          */

proc iml;
   k = 7;
   maxtol = 0.0001;  maxiter = 10000;

   theta1 = 3;    theta2 = 5;           /* user-supplied initial values     */
   maxl = 22*log(5.5) - 4*5.5;          /* user-supplied max log-likelihood */
   start ell;
      l = 22*log(theta) - 4*theta;      /* user-supplied log-likelihood     */
      g = l - maxl + log(k);
   finish;

   iter=0;
   theta = theta1;  run ell;  g1 = g;
   theta = theta2;  run ell;  g2 = g;
   if (g1*g2 > 0) then print "Initial values are improper!!!";
   tol = theta2 - theta1;

   do while (iter <= maxiter & tol > maxtol);
      theta = (theta1 + theta2)/2;
      run ell;
      if (g1*g < 0) then do;  theta2 = theta;  g2 = g;  end;
         else do;             theta1 = theta;  g1 = g;  end;
      iter = iter + 1;
      tol = theta2 - theta1;
   end;

   theta = (theta1 + theta2)/2;
   print iter theta1[format=12.4] theta2[format=12.4] theta[format=12.4];
quit;
```

계산해봅니다. 그 결과

1) $g(\theta)$의 부호가 $g(\theta_1)$의 부호와 다르면, 앞 구간 (θ_1, θ)에

2) 그렇지 않으면, 뒤 구간 (θ, θ_2)에

근이 존재합니다. 이런 식으로 구간의 폭이 충분히 작아질 때까지 구간을 절반씩 좁혀가면서 근에 접근해 가는 방법이 이분법입니다. 일종의 '토끼 몰이'라고 할 수 있습니다. ■



<그림 3> 로그 가능도의 1계 미분 $\eta = \frac{d}{d\theta}\log_e L(\theta)$ 와 접선 및 개선된 θ 값

<표 1>의 프로그램을 보십시오. 4.1절의 포아송 자료에 대하여 구간 $(3,5)$ 에서 근을 찾는 예입니다. 그 결과로 근 3.4990 을 얻습니다. 구간을$(3,5)$ 에서$(6,9)$ 로 바꿔 프로그램을 돌리면 그 때에는 근 8.1483 을 얻습니다. 따라서 1/7 규칙에 의한 구간은 $(3.50, 8.15)$ 입니다. 4.1절의 베르누이 자료에 대하여 <표 1>의 프로그램을 적절히 수정 · 수행함으로써 구간 $(0.471, 0.665)$ 를 얻어 보십시오 (연습문제 4.1).

다음으로 가능도 $L(\theta)$ 가 가장 큰 값이 되도록 하는 파리미터인 최대가능도추정값(mle)을 수치적으로 구하는 뉴튼-라프슨 방법을 설명하겠습니다.

뉴튼-라프슨(Newton-Raphson) 방법

최대가능도추정값(mle) $\hat{\theta}$ 은

$$L(\hat{\theta}) = \max_\theta L(\theta)$$

를 만족해야 합니다. 그러므로 가능도함수 $L(\theta)$ 가 미분가능한 경우

$$\frac{d}{d\theta}\log_e L(\theta) \left(= \frac{L'(\theta)}{L(\theta)}\right) = 0$$

이 되어야 합니다. 따라서

$$g(\theta) = \frac{d}{d\theta}\log_e L(\theta)$$

로 놓고 $g(\theta) = 0$ 의 근(根)을 구하면 됩니다. 그러므로 앞서 설명한 이분법을 활용해도 됩니다. 그러나 뉴튼-라프슨 방법이 이분법보다 더 효율적입니다. 이 방법의 아이디어는 초기값 θ_0 에서 접선(= 함수 $g(\theta)$ 에 대한 선형 근사식)을 구하고 접선의 근으로 θ 값을 개선한다는 것입니다. θ_0 에서 접선의

방정식이

$$\eta = g(\theta_0) + g'(\theta_0)(\theta - \theta_0)$$

입니다. <그림 3>을 참조하세요. 따라서 접선의 근은

$$\theta_1 = \theta_0 - \frac{g(\theta_0)}{g'(\theta_0)},$$

즉

$$\theta_1 = \theta_0 - \frac{\frac{d}{d\theta}\log_e L(\theta_0)}{\frac{d^2}{d\theta^2}\log_e L(\theta_0)} \qquad (4)$$

입니다. 뉴튼-라프슨 방법은 추정값 개선의 정도가 무시될 수 있을 때까지 식 (4)를 반복식으로 씁니다. ■

예로서, 포아송 표본의 가능도 (4)를 최대화하기 위한 뉴튼-라프슨 절차를 살펴보도록 하겠습니다.

$$\log_e L(\theta) = 22\log\theta - 4\theta + \text{constant}$$

이므로

$$g(\theta) = \frac{d}{d\theta}\log_e L(\theta) = \frac{22}{\theta} - 4,$$

$$g'(\theta) = \frac{d^2}{d\theta^2}\log_e L(\theta) = -\frac{22}{\theta^2}$$

입니다. 따라서 초기값을 $\theta_0 = 4$로 두면

$$g(\theta_0) = \frac{d}{d\theta}\log_e L(\theta_0) = \frac{22}{4} - 4 = 1.5,$$

$$g'(\theta_0) = \frac{d^2}{d\theta^2}\log_e L(\theta_0) = -\frac{22}{4^2} = -1.375$$

입니다. 그러므로, 접선의 방정식은

$$\eta = 1.5 + (-1.375)(\theta - 4)$$

가 되며, 식 (4)에 의하여 θ 값(= 접선의 근)이

$$\theta_1 = 4 - \frac{1.5}{-1.375} = 5.09$$

로 개선됩니다. 이것은 초기값 $\theta_0 = 4$에 비교해 우리가 알고 있는 $\hat{\theta} = 5.5$에 근접하는 수치입니다. <그림 3>을 참조해 보십시오.

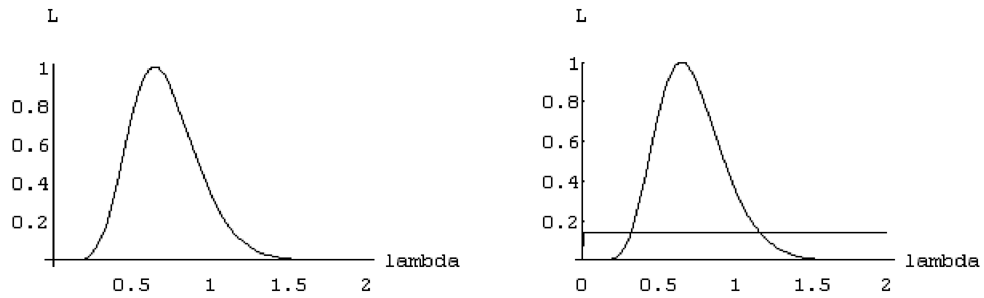
<표 2> 뉴튼-라프슨 방법으로 최대가능도 추정값 구하기

```
/* Newton-Raphson method for MLE computing */
/* mle.iml                                 */

proc iml;
   theta = 4;                            /* user-supplied initial value      */
   maxtol = 0.0001;  maxiter = 10000;

   start ell;
      lik  = 22*log(theta) - 4*theta;    /* user-supplied log-likelihood     */
      lik1 = 22/theta - 4;               /* user-supplied first derivative   */
      lik2 = -22/(theta*theta);          /* user-supplied second derivative  */
   finish;

   iter = 0;92
   tol  = 1;
   do while (iter <= maxiter & tol > maxtol);
      run ell;
      theta1 = theta - lik1/lik2;
      tol = abs(theta1 - theta);
      theta = theta1;
      iter = iter + 1;
   end;
   print iter theta[format=12.4] lik1[format=12.4] lik[format=12.4];
quit;
```

다음 단계에서는 $\theta_0 = 5.09$ 로 하여 θ 에 대한 개선값을 계산하게 됩니다. <표 2>의 프로그램을 보십시오.

※ 서브루틴 ell에 로그 가능도 함수 lik, 로그 가능도의 1계 미분 lik1 및 2계 미분 lik2를 지정하여야 합니다. 또한 적절한 초기값 설정도 중요합니다.

물론, 포아송 표본의 가능도는 미분으로 쉽게 최대점을 찾을 수 있기 때문에 뉴튼-라프슨 방법과 같은 수치적인 방법을 쓸 필요가 없습니다. 그렇지만 어떤 예의 경우는 수치적 방법이 아닌 순수한 해석적인 방법으로는 도저히 풀리지 않는답니다. 그런 한 예를 보도록 하지요.



<그림 4> 이산화 지수 표본의 정규화 가능도

전자부품의 수명 $t_1, \cdots, t_n$에 대하여 지수분포 Exponential(θ)로 확률모형화한다고 합시다. 이런 관측값들은 연속적인 실수값이여야 합니다. 그런데 부품이 작동하고 있는가 그렇지 않은가가 실제로는 정수 시각 $1, 2, 3, \cdots$ 에서 점검된다고 합시다. 따라서 실제의 데이터는 $t_1, \cdots, t_n$을 이산화한 $x_1, \cdots, x_n$입니다. 여기서 $t_1, \cdots, t_n$과 $x_1, \cdots, x_n$ 사이의 관계는 다음과 같습니다 (예컨대 $t_1 = 2.4$이면 $x_1 = 3$이므로).

$$x_1 = [t_1] + 1, \cdots, x_n = [t_n] + 1.$$

그러므로 정수값 확률변수 X는 다음과 같은 확률함수를 갖습니다.

$$P\{X = x\} = P\{x-1 < T \le x\} = \int_{x-1}^{x} \frac{1}{\theta} e^{-\frac{t}{\theta}} dt$$

$$= e^{-\frac{x-1}{\theta}} - e^{-\frac{x}{\theta}}, \quad x = 1, 2, 3, \cdots .$$

따라서 가능도 $L(\theta)$가 다음과 같게 됩니다.

$$L(\theta) = \prod_{i=1}^{n} \left\{ e^{-\frac{x_i - 1}{\theta}} - e^{-\frac{x_i}{\theta}} \right\}.$$

이산화 지수 표본 자료 ($x = 1,1,1,1,1,1,2,3,4,6$; $n = 10$)에 대한 가능도인 <그림 4>를 보십시오.

편의상 $\frac{1}{\theta} = \lambda$, $p(x;\lambda) = e^{-\lambda(x-1)} - e^{-\lambda x}$로 놓으면

$$\frac{d}{d\lambda} p(x;\lambda) = -(x-1) e^{-\lambda(x-1)} + x e^{-\lambda x},$$

$$\frac{d^2}{d\lambda^2} p(x;\lambda) = (x-1)^2 e^{-\lambda(x-1)} - x^2 e^{-\lambda x}$$

가 됩니다. 이것을 이용하여 로그 가능도, 로그 가능도의 1계 미분 · 2계 미분이 다음과 같이 표현 가능합니다.

$$\log_e L(\lambda) = \sum_{i=1}^{n} \log_e p(x_i;\lambda),$$

$$\frac{d}{d\lambda}\log_e L(\lambda) = \sum_{i=1}^{n} \frac{\frac{d}{d\lambda}p(x_i;\lambda)}{p(x_i;\lambda)},$$

$$\frac{d^2}{d\lambda^2}\log_e L(\lambda) = \sum_{i=1}^{n}\left[\frac{\frac{d^2}{d\lambda^2}p(x_i;\lambda)}{p(x_i;\lambda)} - \left\{\frac{\frac{d}{d\lambda}p(x_i;\lambda)}{p(x_i;\lambda)}\right\}^2\right].$$

초기값을 $\lambda_0 = 1$로 놓고 <표 2>의 프로그램을 수정하여 mle $\hat{\lambda}$을 구하니 0.6466이 나오는군요. 즉 해당하는 지수분포 Exponential(λ^{-1})의 평균은 0.6466^{-1}, 즉 1.547입니다. 자료값들의 평균 2.1에 비하면 0.553이 작군요 (참고: 올림 값과 실제 값의 차이는 평균적으로 0.5임). 이 정도면 뉴튼-라프슨 방법과 최대가능도추정값이 대견하지 않습니까? 참고로, 이산화(discretized) 지수 표본 자료의 가능도에 1/7 규칙을 적용하여 파라미터 λ의 구간을 구한 결과는$(0.320, 1.15)$입니다 (연습문제 4.4).

4.3 가능도의 여러 사례들

앞의 두 절에서 가능도에 관한 전형적인 사례들을 보았습니다. 파리미터가 하나이고 가능도가 단봉(單峰, unimode)의 매끄러운 형태인 것이 특징이지요. 이 절에서는 이와 다른 사례들을 접하게 될 것입니다.

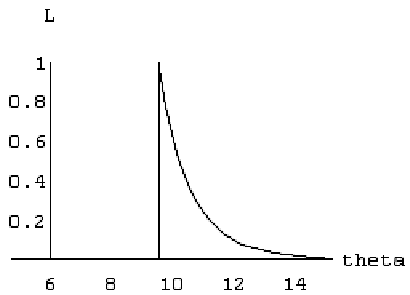
$x_1, \cdots, x_n$을 균일분포 Uniform$(0, \theta)$로 모형화 한다고 합시다. 그러면

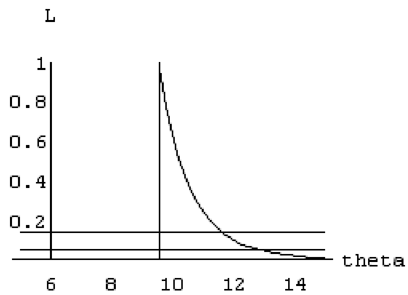
$$L(\theta) = \left(\frac{1}{\theta}\right)^n, \quad 0 \le x_1, \cdots, x_n \le \theta$$

가 됩니다. 더 정확히 표현하자면

$$L(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & \theta \ge \max(x_1, \cdots, x_n) \\ 0, & \text{otherwise} \end{cases}$$

입니다. 이 함수는 $\hat{\theta} = \max(x_1, \cdots, x_n)$에서 최대값을 갖습니다. 그런데 이 점에서 $L(\theta)$가 불연속이라는 점에서 특이합니다. 바로 비정칙적(非正則的, irregular)인 경우입니다.





<그림 5> 균일분포 자료의 정규화 가능도

<그림 5>는 자료

$$4.95, 9.37, 0.49, 4.64, 9.55, 5.31, 7.43, 7.52, 4.85, 0.25 \quad (n = 10)$$

에 대한 정규화 가능도를 그린 것입니다. $\hat{\theta} = 9.55$ 이고 이 점의 왼편은 낭떠러지인 것을 볼 수 있습니다. k분의 1 규칙을 적용하기 위하여 $k = 7$과 $k = 20$에 수평선을 그은 것이 <그림 5>의 오른 쪽 그림입니다. 통상적으로는 k를 7로 하였었는데 이 사례처럼 정칙적(regular)이지 않은 경우에는 좋지 않습니다. 나중에 밝혀지겠지만, 이 경우엔 $k = 20$ 이 좋습니다. 그렇게 구한 θ의 구간은 $(9.55, 12.89)$ 입니다.

두 번째 사례는 정규분포에 관한 것입니다. $x_1, \cdots, x_n$을 정규분포 $N(\mu, \sigma^2)$로 모형화한다고 합시다. 그리고 μ와 σ^2가 모두 미지(未知)라고 합시다. 그러므로 이 사례는 파라미터가 2개라는 점에서 이제까지의 사례들과는 다릅니다. 2개의 파라미터를 파라미터 벡터

$$\boldsymbol{\theta} = (\theta_1, \theta_2), \quad 즉\ \theta_1 = \mu,\ \theta_2 = \sigma^2$$

으로 표기하기로 하겠습니다. 그러면

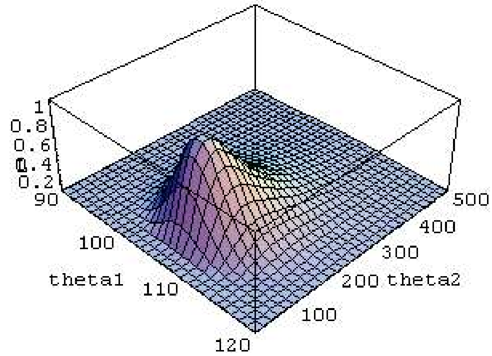
$$L(\theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi}}\right) \theta_2^{-n/2} \exp\left\{-\sum_{i=1}^{n} (x_i - \theta_1)^2 / (2\theta_2)\right\}, \quad \theta_2 > 0$$

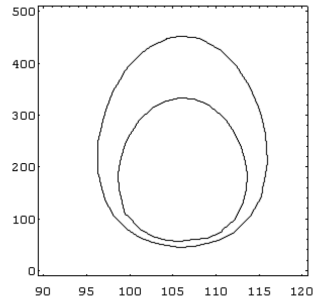
으로 표현됩니다. 따라서

$$l(\theta_1, \theta_2) = -\frac{n}{2} log_e \theta_2 - \sum_{i=1}^{n} (x_i - \theta_1)^2 / (2\theta_2) + \text{constant}, \quad \theta_2 > 0,$$

$$\frac{\partial}{\partial \theta_1} l(\theta_1, \theta_2) = \sum_{i=1}^{n} (x_i - \theta_1) / \theta_2,$$

$$\frac{\partial}{\partial \theta_2} l(\theta_1, \theta_2) = -\frac{n}{2\theta_2} + \frac{\sum_{i=1}^{n} (x_i - \theta_1)^2}{2\theta_2^2}$$





<그림 6> 정규분포 표본의 가능도 (정규화)

이 됩니다. 그러므로

$$\frac{\partial}{\partial\theta_1} l(\theta_1, \theta_2) = 0 \;\Rightarrow\; \hat{\theta}_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \;,$$

$$\frac{\partial}{\partial\theta_1} l(\theta_1, \theta_2) = 0,\; \frac{\partial}{\partial\theta_2} l(\theta_1, \theta_2) = 0 \;\Rightarrow\; \hat{\theta}_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

입니다. <그림 6>의 왼쪽 3D 플롯은 10개의 관측값 자료

$$105, 86, 107, 93, 122, 101, 113, 118, 103, 112 \; (n = 10)$$

에 대한 정규화 가능도입니다($\hat{\theta}_1 = 106, \hat{\theta}_2 = 109$).

<그림 6>의 오른쪽 그림은 k분의 1 영역인데 안쪽 작은 달걀이 $k = 7$인 경우이고 바깥쪽 큰 달걀은 $k = 20$인 경우입니다. 파라미터가 2개인 경우에는 파라미터가 1개인 경우에 비해 k를 더 크게 해야 합니다. 나중에 이유를 밝히겠지만 앞으로 파라미터가 2개이고 정칙적인 경우에는 $k = 20$을 사용하여, 즉 1/20 규칙에 의하여, (θ_1, θ_2) 영역을 구하게 됩니다.

파라미터가 2개인 경우, (θ_1, θ_2)에 대한 mle를 찾기 위해서 일반적으로는 수치적인 방법까지 동원하여야 합니다. 정규분포의 가능도의 경우에는 운 좋게도 해석적인 방법으로 최대점을 찾을 수 있었지만, 인생 살다보면 갠 날만 있는 것은 아니잖습니까? 앞 절에서 다루었던 뉴튼-라프슨 방법을 확장하여 보도록 하겠습니다.

뉴튼-라프슨(Newton-Raphson) 방법 : 파리미터가 2개인 경우

정칙적인 경우 (가능도가 어디서나 미분 가능하고 최적점이 파라미터 공간의 내부에 있는 경우), mle $\hat{\theta}_1$과 $\hat{\theta}_2$을 구하려면 로그 가능도 $l(\theta_1, \theta_2)$에 대한 기울기 벡터(gradient vector)

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \dfrac{\partial l}{\partial \theta_1} \\ \\ \dfrac{\partial l}{\partial \theta_2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{즉} \quad \frac{\partial l}{\partial \theta_1} = 0, \ \frac{\partial l}{\partial \theta_2} = 0$$

이여야 합니다. 그런데 초기 점$(\theta_{10}, \theta_{20})$에서 $\dfrac{\partial l}{\partial \boldsymbol{\theta}}$에 대한 선형근사는

$$\begin{pmatrix} \dfrac{\partial l}{\partial \theta_1} \\ \dfrac{\partial l}{\partial \theta_2} \end{pmatrix} \simeq \begin{pmatrix} \dfrac{\partial l_0}{\partial \theta_1} \\ \dfrac{\partial l_0}{\partial \theta_2} \end{pmatrix} + \begin{pmatrix} \dfrac{\partial l_0^2}{\partial^2 \theta_1} & \dfrac{\partial l_0^2}{\partial \theta_1 \, \partial \theta_2} \\ \dfrac{\partial l_0^2}{\partial \theta_1 \, \partial \theta_2} & \dfrac{\partial l_0^2}{\partial^2 \theta_2} \end{pmatrix} \begin{pmatrix} \theta_1 - \theta_{10} \\ \theta_2 - \theta_{20} \end{pmatrix}$$

↑ 헤시안(Hessian) 행렬

로 표현됩니다. 따라서 좌변을 0으로 놓으면

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \theta_{10} \\ \theta_{20} \end{pmatrix} - \begin{pmatrix} \dfrac{\partial l_0^2}{\partial^2 \theta_1} & \dfrac{\partial l_0^2}{\partial \theta_1 \, \partial \theta_2} \\ \dfrac{\partial l_0^2}{\partial \theta_1 \, \partial \theta_2} & \dfrac{\partial l_0^2}{\partial^2 \theta_2} \end{pmatrix}^{-1} \begin{pmatrix} \dfrac{\partial l_0}{\partial \theta_1} \\ \dfrac{\partial l_0}{\partial \theta_2} \end{pmatrix}$$

가 유도됩니다. 바로 이 식을 반복적으로 사용하자는 것이지요. 파라미터가 1개의 경우의 식 (4)와의 유사성을 확인해 보십시오. ■

어느 특정 처리를 받은 실험 쥐의 수명자료

$$1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23 \ (n = 21)$$

을 와이블 분포 Weibull(γ, β)로 모형화하는 사례를 생각해 보겠습니다 (연습문제 1.2 참조). 표기를 일치시키기 위하여 $\theta_1 = \gamma, \theta_2 = \beta$라고 하겠습니다. Weibull$(\theta_1, \theta_2)$의 확률밀도는

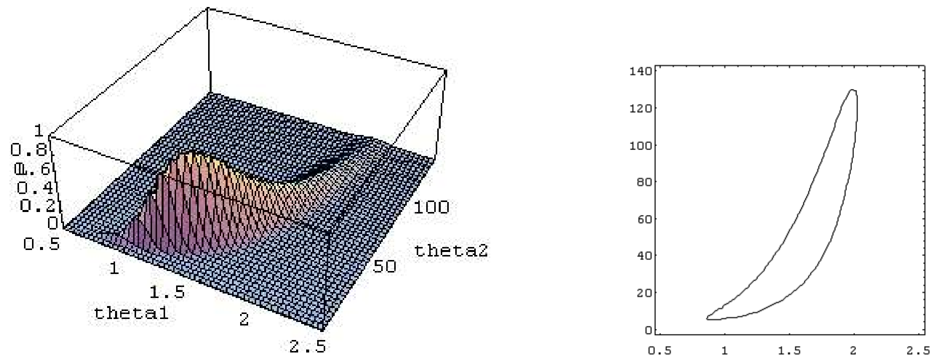
$$f(x; \theta_1, \theta_2) = \frac{\theta_1}{\theta_2} x^{\theta_1 - 1} \exp\left(-\frac{1}{\theta_2} x^{\theta_1}\right), \ \theta_1 > 0, \ \theta_2 > 0$$

입니다. 그러므로 로그 가능도가

$$l(\theta_1, \theta_2) = n \log_e \theta_1 - n \log_e \theta_2 + (\theta_1 - 1) \sum_{i=1}^{n} \log_e x_i - \frac{1}{\theta_2} \left(\sum_{i=1}^{n} x_i^{\theta_1} \right)$$

<표 3> 2개 파라미터의 최대가능도 추정 : 뉴튼-라프슨 방법

```
/* Newton-Raphson method for two parameters */
/* mle2.iml                                 */

proc iml;
   theta1 = 1;  theta2 = 8.67;              /* user-supplied initial values */
   theta = theta1//theta2;
   maxtol = 0.0001;  maxiter = 10000;

   start ell;
      x={1 1 2 2 3 4 4 5 5 8 8 8 8 11 11 12 12 15 17 22 23};
      n = 21;
      temp0 = exp(theta1*log(x));
      temp1 = exp(theta1*log(x))#log(x);
      temp2 = exp(theta1*log(x))#log(x)#log(x);
      /* user-supplied log-likelihood function */
         lik = n*log(theta1/theta2) + (theta1-1)*sum(log(x))
               - sum(temp0)/theta2;
      /* user-supplied first derivatives */
         lik1 = n/theta1 + sum(log(x)) - sum(temp1)/theta2;
         lik2 =-n/theta2 + sum(temp0)/(theta2*theta2);
      /* user-supplied second derivatives */
         lik11 = -n/(theta1*theta1) - sum(temp2)/theta2;
         lik22 = n/(theta2*theta2) - 2*sum(temp0)/(theta2*theta2*theta2);
         lik12 = sum(temp1)/(theta2*theta2);
      /* gradient and Hessian matrix */
         grad = lik1//lik2;
         Hess = (lik11||lik12)//(lik12||lik22);
   finish;

   iter = 0;
   tol  = 1;
   do while (iter <= maxiter & tol > maxtol);
      run ell;
      update = theta - inv(Hess)*grad;
      tol = ssq(update - theta);
      theta = update;
      theta1 = theta[1];
      theta2 = theta[2];
      iter = iter + 1;
   end;
   print iter theta[format=12.4] grad[format=12.4] lik[format=12.4];
quit;
```



<그림 7> 와이블 표본 자료의 가능도 (정규화)

로 표현됩니다. 그리고 1계 미분과 2계 미분이 각각

$$\frac{\partial l}{\partial \theta_1} = \frac{n}{\theta_1} + \sum_{i=1}^{n} \log_e x_i - \frac{1}{\theta_2}\left(\sum_{i=1}^{n} x_i^{\theta_1} \log_e x_i\right),$$

$$\frac{\partial l}{\partial \theta_2} = -\frac{n}{\theta_2} + \frac{1}{\theta_2^2}\left(\sum_{i=1}^{n} x_i^{\theta_1}\right)$$

과

$$\frac{\partial^2 l}{\partial \theta_1^2} = -\frac{n}{\theta_1^2} - \frac{1}{\theta_2}\left\{\sum_{i=1}^{n} x_i^{\theta_1} (\log_e x_i)^2\right\},$$

$$\frac{\partial^2 l}{\partial \theta_2^2} = \frac{n}{\theta_2^2} - \frac{2}{\theta_2^3}\left(\sum_{i=1}^{n} x_i^{\theta_1}\right)$$

$$\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} = \frac{1}{\theta_2^2}\left(\sum_{i=1}^{n} x_i^{\theta_1} \log_e x_i\right)$$

가 됩니다. <표 3>은 이 방법을 구현한 프로그램입니다.

<표 3>을 돌린 결과 불과 7번의 반복수행으로 $\hat{\theta}_1 = 1.371$, $\hat{\theta}_2 = 21.82$ 가 나왔습니다. <그림 7>의 왼쪽 그림을 보십시오. 그리고 1/20 규칙에 의하여 (θ_1, θ_2) 영역을 구한 것이 <그림 7>의 오른 쪽 그림입니다.

그림을 감상해 보세요. 공룡이 모래밭에 머리를 처 박고 있는 모습이 아닙니까?

4.4 충분성 (Sufficiency)

가능도 $L(\theta)$는 θ의 함수입니다만 표본자료 $x_1, \cdots, x_n$에 따라 다르게 정의됩니다. 그래서

$$L(\theta) = L(\theta; x_1, \cdots, x_n) \ (= \prod_{i=1}^{n} f(x_i; \theta),\ \text{임의표본의 경우})$$

으로 쓸 수 있습니다. 즉, 가능도 $L(\theta; x_1, \cdots, x_n)$은 θ의 함수이되 표본 자료값에 따라 달리 정의된다는 것을 강조한 표기입니다. 예로서, 4.1절에서

$$x_1, x_2, x_3, x_4 \ (= 3, 10, 5, 4) \quad \sim \quad \text{iid} \quad \text{Poisson}(\theta)$$

인 경우를 다루었었는데, 이 때 표본자료 x_1, x_2, x_3, x_4가 4, 12, 8, 9 이면 가능도는

$$L(\theta; 4,12,8,9) = \frac{\theta^{33} e^{-4\theta}}{4!12!8!9},\ \theta > 0 \qquad (5)$$

이 됩니다. 이 함수는 $\hat{\theta} = 33/4 \ (= 8.25)$에서 최대값을 갖는 등 <그림 1>의 가능도와는 다릅니다. 만약 표본자료가 6, 10, 6, 11 이면 가능도는

$$L(\theta; 6,10,6,11) = \frac{\theta^{33} e^{-4\theta}}{6!10!6!11!},\ \theta > 0 \qquad (6)$$

가 됩니다만, (5)와 (6)의 가능도는 사실상 동일합니다. 왜냐하면 정규화를 하는 경우 상수인 분모항은 없어지기 때문입니다. 일반적으로 포아송 임의표본의 경우, 가능도

$$L(\theta; x_1, \cdots, x_n) = \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!},\ \theta > 0$$

은 $\sum_{i=1}^{n} x_i$에 따라 달라질 뿐입니다. 다시 말하여 표본자료가 다르더라도 합만 같으면 가능도는 사실상 동일합니다. 그런데 표본 자체는 n-변량인 $(x_1, \cdots, x_n)$이지만 합은 단(單)변량 $t \ (= \sum_{i=1}^{n} x_i)$입니다. 훨씬 간단한 것이지요. 그런데

$$T \ (= \sum_{i=1}^{n} X_i) \quad \sim \quad \text{Poisson}(n\theta)$$

를 따릅니다. 그러므로, 조건부 확률밀도는

$$f_{X_1, \cdots, X_n \mid T}(x_1, \cdots, x_n \mid t; \theta) = \frac{f_{X_1, \cdots, X_n}(x_1, \cdots, x_n; \theta)}{f_T(t; \theta)}$$

$$= \frac{\theta^{\sum_{i=1}^{n} x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!} \Bigg/ \frac{(n\theta)^t e^{-n\theta}}{t!} = \frac{t!}{\prod_{i=1}^{n} x_i!}\left(\frac{1}{n}\right)^t, \quad t = \sum_{i=1}^{n} x_i$$

로서 θ와 무관합니다 (이 분포는 시행의 수가 t이고 항의 수가 n이며 각 항의 확률이 $1/n$로 동일한 다항분포(multinomial distribution)입니다). 그러므로 θ에 대하여 표본 데이터 $x_1, \cdots, x_n$이 갖고 있는 '정보'는 모두 $T = t$ 안에 있다고 할 수 있습니다.

<u>충분통계량</u>(sufficient statistic) : 정의

$t = T(x_1, \cdots, x_n)$이 주어졌을 때, $(x_1, \cdots, x_n)$의 조건부 확률밀도

$$f_{X_1, \cdots, X_n \mid T}(x_1, \cdots, x_n \mid t; \theta) = \frac{f_{X_1, \cdots, X_n}(x_1, \cdots, x_n; \theta)}{f_T(t; \theta)}$$

가 θ에 무관하면, $T = T(X_1, \cdots, X_n)$을 파라미터 θ를 위한 충분통계량이라고 합니다. ■

확률모형과 임의표본이 충분통계량 T를 갖는 경우, 임의표본으로부터의 가능도 $L(\theta; x_1, \cdots, x_n)$은 통계량 T로부터의 가능도와 사실상 동일하게 됩니다. 왜냐하면

$$f_{X_1, \cdots, X_n}(x_1, \cdots, x_n; \theta) = f_{X_1, \cdots, X_n \mid T}(x_1, \cdots, x_n \mid t) \cdot f_T(t; \theta) \propto f_T(t; \theta),$$

즉 θ의 함수로서 비례하기 때문이지요. 따라서, 넓은 의미에서는 θ에 대하여 n-변량 $(x_1, \cdots, x_n)$이 갖는 '정보'와 1-변량 t가 갖는 '정보'는 동등합니다. 이것은 충분통계량 T에 의하여 상당한 데이터 축소(data reduction)가 가능함을 말합니다. 예컨대 $n = 1{,}000{,}000$이라고 해보세요. 백만개의 자료값 $x_1, \cdots, x_{1000000}$을 1개 실수값인 t로 압축한 셈이지 않습니까?

순전히 상상입니다만, 여러분이 게임 창출자 A의 입장에 있다고 하고 충분통계량이 어떤 역할을 하는지 봅시다.

<u>통계 게임</u> : 1) 게임 창출자 A는 확률모형 $f(x; \theta)$로부터 임의표본 $x_1, \cdots, x_n$을 발생시켜 플레이어 B에게 줍니다. 이 때, θ 값은 A 자신만 알고 있습니다.

2) 표본 데이터를 바탕으로 B는 θ에 대한 추측을 합니다. B는 θ 값은 모르지만 $x_1, \cdots, x_n$이 확률모형 $f(x; \theta)$로부터의 임의표본이라는 것을 압니다.

3) 추측이 정확할수록 B는 A로부터 큰 상(또는 작은 벌)을 받습니다.

만약, 이 확률모형과 임의표본이 충분통계량 T를 갖는다면, 게임의 단계 1을 다음과 같이 2개의 부(副)단계로 분할할 수 있을 것입니다.

1-1) 확률모형 $f_T(t;\theta)$ 로부터 1개의 자료값 t 를 만들어냅니다.

1-2) 확률밀도 $f_{X_1,\cdots,X_n \mid T}(x_1,\cdots,x_n \mid t)$ 로부터 n-변량 $(x_1,\cdots,x_n)$ 을 발생시켜 B에게 줍니다.

그러므로 θ 와 무관한 부단계 1-2는 B의 일을 괜스레 복잡하게 만드는 역할을 합니다. 따라서 “똑똑한” B가 우선적으로 해야 할 일은

복잡한 n-변량 $(x_1,\cdots,x_n)$ 을 1-변량 t 로 단순화 !

하는 일입니다. 즉 데이터의 복잡도를 최대한 줄이는 작업이라고 하겠습니다.

충분통계량에 관한 예제를 하나 더 보겠습니다. 균일분포 Uniform$(0,\theta)$ 로부터의 임의표본 $x_1,\cdots,x_n$ 에서 충분통계량은 무엇일까요? 가능도 $L(\theta;x_1,\cdots,x_n)$ 이 무엇이었나 생각해 보십시오 (4.3절 참조). 가능도가 $T \equiv \max(X_1,\cdots,X_n)$ 에 따라 달리 정의되었던 것을 기억할 수 있을 겁니다. 직관적으로, T가 충분통계량일 것이라는 감(感)이 옵니다. 그것을 증명해 봅시다. 어렵지 않게

$$f_T(t;\theta) = \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1}, \quad 0 \le t \le \theta \tag{7}$$

라는 것을 알 수 있습니다. 따라서

$$\begin{aligned} f_{X_1,\cdots,X_n \mid T}(x_1,\cdots,x_n \mid t;\theta) &= \frac{f_{X_1,\cdots,X_n}(x_1,\cdots,x_n;\theta)}{f_T(t;\theta)} \\ &= \left(\frac{1}{\theta}\right)^n \Big/ \left\{\frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1}\right\} = \frac{1}{n\ t^{n-1}}, \quad 0 \le x_1,\cdots,x_n \le t \end{aligned} \tag{8}$$

입니다. 보다시피 조건부 확률밀도가 θ 에 무관하므로 확실하게 T가 충분통계량입니다 (즉, 증명을 마친 것입니다). 이 사례에서, 통계 게임의 부단계 1-1은 (7)의 확률밀도에 따라 자료값 t 가 생성되는 것이고 부단계 1-2는 (8)의 확률밀도에 따라 n 개의 자료값 $x_1,\cdots,x_n$ 이 파생되는 것을 의미합니다. 뒷 단계를 면밀히 볼까요?

i) Uniform$(0,t)$ 로부터 $n-1$ 개의 임의표본 $u_1,\cdots,u_{n-1}$ 을 만듭니다.
따라서 이 때의 결합밀도는 $\left(\frac{1}{t}\right)^{n-1}$ 입니다.

ii) $u_1,\cdots,u_{n-1}$ 의 순서열의 앞 · 뒤, 또는 사이 공간 하나를 골라 t 를 끼워 넣음으로써 $x_1,\cdots,x_n$ 이 파생된다고 볼 수 있습니다. 이때 가능한 경우의 수는 n 입니다. 따라서 확률 $\frac{1}{n}$ 의 사건이 발생합니다.

iii) 확률밀도 $\left(\frac{1}{t}\right)^{n-1}$ 과 확률 $\frac{1}{n}$ 을 곱하여 (8)이라고 할 수 있습니다.

충분통계량을 찾는 일은 앞의 예에서 보았듯이 ① 가능도 함수 $L(\theta; x_1, \cdots, x_n)$ 이 어느 통계량을 통하여 표현되는가를 살펴보고 ② 그 통계량이 충분성의 정의를 충족시키는지를 검토하는 방법이 정통적이라고 하겠습니다. 그러나 다음 정리가 보여주듯 ①의 단계로 충분합니다.

<u>인수화 정리</u>(factorization theorem) :

가능도 $L(\theta; x_1, \cdots, x_n)$ 가

$$L(\theta; x_1, \cdots, x_n) = g(T(x_1, \cdots, x_n); \theta) \cdot h(x_1, \cdots, x_n) \tag{9}$$

의 형태로 인수화되면, $T = T(X_1, \cdots, X_n)$ 은 파라미터 θ 를 위한 충분통계량입니다. 그리고 그 역(逆)도 성립합니다.

증명은 간단합니다. $X_1, \cdots, X_n$ 이 연속형 변수인 경우를 생각하기로 합니다. (9)를 가정하면

$$
\begin{aligned}
f_{X_1, \cdots, X_n \mid T}(x_1, \cdots, x_n \mid t; \theta) &= \frac{f_{X_1, \cdots, X_n}(x_1, \cdots, x_n; \theta)}{f_T(t; \theta)} \\
&= \frac{g(T(x_1, \cdots, x_n); \theta) \cdot h(x_1, \cdots, x_n)}{f_T(t; \theta)} \\
&= \frac{g(T(x_1, \cdots, x_n); \theta) \cdot h(x_1, \cdots, x_n)}{\int_{A_t} g(T(x_1, \cdots, x_n); \theta) \cdot h(x_1, \cdots, x_n)\, dx_1 \cdots dx_n} \\
&\qquad (\text{여기서 } A_t = \{(x_1, \cdots, x_n) \mid T(x_1, \cdots, x_n) = t\}) \\
&= \frac{g(T(x_1, \cdots, x_n); \theta) \cdot h(x_1, \cdots, x_n)}{g(T(x_1, \cdots, x_n); \theta) \int_{A_t} h(x_1, \cdots, x_n)\, dx_1 \cdots dx_n} \\
&= \frac{h(x_1, \cdots, x_n)}{\int_{A_t} h(x_1, \cdots, x_n)\, dx_1 \cdots dx_n}
\end{aligned}
$$

로 파라미터 θ 와 무관하다는 사실이 유도됩니다. $X_1, \cdots, X_n$ 이 이산형 변수인 경우엔 위에서 $\int$ 을 $\sum$ 로 대치하면 됩니다.

역방향의 증명은 충분통계량의 정의로부터 자명(自明)하게 나옵니다. ■

응용 예를 하나 들어보기로 합니다. $x_1, \cdots, x_n$에 대하여 Exponential(θ) 분포를 가정하기로 합시다. 그러면 가능도 $L(\theta)$가

$$L(\theta\,; x_1, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta} \sum_{i=1}^{n} x_i\right),\ \theta > 0$$

로 표현됩니다. 그러므로

$$T(X_1, \cdots, X_n) = \sum_{i=1}^{n} X_i$$

로, 함수 $g(t\,;\theta)$와 $h(x_1,\cdots,x_n)$을 각각

$$g(t\,;\theta) = \frac{1}{\theta^n} \exp\left(-\frac{t}{\theta}\right), \quad t \ge 0,$$

$$h(x_1,\cdots,x_n) = 1, \quad x_1 \ge 0, \cdots, x_n \ge 0$$

로 잡으면 가능도 $L(\theta)$가 식 (9)의 형태로 인수화됩니다. 따라서 $T = \sum_{i=1}^{n} X_i$가 충분통계량입니다.

4.5 충분통계량의 여러 사례들

표본자료 $x_1, \cdots, x_n$에 대하여 Uniform(θ_1, θ_2) 분포를 가정한다고 합시다. 그러면 가능도 $L(\theta_1, \theta_2)$는

$$L(\theta_1,\theta_2) = \left(\frac{1}{\theta_2 - \theta_1}\right)^n, \quad \theta_1 \le x_1, \cdots, x_n \le \theta_2$$

가 됩니다. 이 가능도에서 $x_1, \cdots, x_n$에 연동된 부분은 제약식

$$T_1 \equiv \min(X_1, \cdots, X_n) \ge \theta_1, \quad T_2 \equiv \max(X_1, \cdots, X_n) \le \theta_2$$

에 있습니다. 그러므로 T_1과 T_2가 무언가 핵심적인 통계량이라는 것을 직감적으로 알 수 있습니다.

<u>결합충분통계량</u>(jointly sufficient statistics) : 정의

$t_1 = T_1(x_1, \cdots, x_n)$과 $t_2 = T_2(x_1, \cdots, x_n)$가 주어졌을 때, $(x_1, \cdots, x_n)$의 조건부 확률밀도

$$f_{X_1, \cdots, X_n \mid T_1, T_2}(x_1, \cdots, x_n \mid t_1, t_2\,; \theta_1, \cdots, \theta_k) = \frac{f_{X_1, \cdots, X_n}(x_1, \cdots, x_n\,; \theta_1, \cdots, \theta_k)}{f_{T_1, T_2}(t_1, t_2\,; \theta_1, \cdots, \theta_k)}$$

가 $\theta_1, \cdots, \theta_k$에 무관하면, $T_1 = T_1(X_1, \cdots, X_n)$과 $T_2 = T_2(X_1, \cdots, X_n)$을 파라미터 $\theta_1, \cdots, \theta_k$를 위한 결합충분통계량이라고 합니다. ■

앞의 Uniform(θ_1, θ_2) 표본자료의 사례를 계속하겠습니다. T_1과 T_2가 정말 결합 충분한지 검사해 봅시다. (T_1, T_2)의 결합확률밀도 $f_{T_1,T_2}(t_1, t_2\,; \theta_1, \theta_2)$는

$$f_{T_1,T_2}(t_1, t_2\,; \theta_1, \theta_2)\, dt_1\, dt_2 = n\,(n-1)\,\frac{dt_1}{\theta_2-\theta_1}\left(\frac{t_2-t_1}{\theta_2-\theta_1}\right)^{n-2}\frac{dt_2}{\theta_2-\theta_1}$$

로부터 (왜냐하면 무한소구간(t_1, t_1+dt_1)에 1개의 관측, 구간(t_1, t_2)에 $n-2$개의 관측이 떨어져야 하고, 무한소구간(t_2, t_2+dt_2)에 마지막 1개의 관측이 있어야 하므로) 다음과 같이 얻어집니다.

$$f_{T_1,T_2}(t_1, t_2\,; \theta_1, \theta_2) = n\,(n-1)\,\frac{1}{(\theta_2-\theta_1)^2}\left(\frac{t_2-t_1}{\theta_2-\theta_1}\right)^{n-2}, \quad \theta_1 \le t_1 \le t_2 \le \theta_2.$$

따라서

$$f_{X_1,\cdots,X_n\,|\,T_1,T_2}(x_1, \cdots, x_n \mid t_1, t_2\,; \theta_1, \theta_2) = \frac{f_{X_1,\cdots,X_n}(x_1, \cdots, x_n\,; \theta_1, \theta_2)}{f_{T_1,T_2}(t_1, t_2\,; \theta_1, \theta_2)}$$

$$= \frac{\dfrac{1}{(\theta_2-\theta_1)^n}}{n\,(n-1)\,\dfrac{1}{(\theta_2-\theta_1)^2}\left(\dfrac{t_2-t_1}{\theta_2-\theta_1}\right)^{n-2}}$$

$$= \frac{1}{n\,(n-1)\,(t_2-t_1)^{n-2}}, \quad t_1 \le x_1, \cdots, x_n \le t_2$$

로 (θ_1, θ_2)와 무관합니다. 그러므로 T_1과 T_2가 결합충분통계량입니다.

<u>확장 인수화 정리</u>(extended factorization theorem) : k개 파라미터의 경우.

가능도 $L(\theta_1, \cdots, \theta_k\,; x_1, \cdots, x_n)$가

$$L(\theta_1, \cdots, \theta_k\,; x_1, \cdots, x_n)$$
$$= g\,(T_1(x_1, \cdots, x_n), \cdots, T_l(x_1, \cdots, x_n)\,; \theta_1, \cdots, \theta_k) \cdot h\,(x_1, \cdots, x_n)$$

의 형태로 인수화되면, $T_1 = T_1(X_1, \cdots, X_n), \cdots, T_l = T_l(X_1, \cdots, X_n)$은 파라미터 $\theta_1, \cdots, \theta_k$를 위한 결합충분통계량입니다. 그리고 그 역(逆)도 성립합니다. 통상적으로 $k = l$입니다 (항상 그런 것은 아닙니다).

증명은 앞의 1개 파라미터를 위한 인수화 정리의 증명과 같습니다. ■

앞의 Uniform(θ_1,θ_2) 사례에선

$$L(\theta_1,\theta_2) = \left(\frac{1}{\theta_2-\theta_1}\right)^n, \quad \theta_1 \le t_1 \le t_2 \le \theta_2$$

이므로, 확장인수화정리에 의해 (T_1, T_2)가 결합충분통계량입니다.

두 번째 사례는 표본자료 $x_1, \cdots, x_n$에 대하여 N(θ_1,θ_2) 분포를 가정한 경우입니다 (여기서 $\theta_1 = \mu, \theta_2 = \sigma^2$). 이 때에는

$$\begin{aligned} L(\theta_1,\theta_2) &= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\theta_2}} \exp\left\{-\frac{(x_i-\theta_1)^2}{2\theta_2}\right\}\right] \\ &= \text{constant} \cdot \theta_2^{-n/2} \exp\left\{-\sum_{i=1}^n \frac{(x_i-\theta_1)^2}{2\theta_2}\right\} \\ &= \text{constant} \cdot \theta_2^{-n/2} \exp\left\{-\sum_{i=1}^n \frac{(x_i-\bar{x})^2}{2\theta_2} - n\frac{(\bar{x}-\theta_1)^2}{2\theta_2}\right\} \end{aligned}$$

이므로, 확장인수화정리에 의해

$$(\bar{X}, \sum_{i=1}^n (X_i-\bar{X})^2), \quad \text{즉} \ (\bar{X}, S^2)$$

이 결합충분통계량입니다 (여기서 $S^2 = \sum_{i=1}^n (X_i-\bar{X})^2/(n-1)$, 표본분산). $\bar{X}$는 정규분포 N$(\theta_1, \theta_2/n)$를 따르고, $\bar{X}$와 독립적으로 $\sum_{i=1}^n (X_i-\bar{X})^2$는 $\theta_2 \cdot \chi^2(n-1)$ 분포를 따릅니다 (2.3절 참조).

세번째 사례로 $x_1, \cdots, x_n$에 대하여 Exponential(θ_1,θ_2) 분포를 가정한 경우를 생각합시다. 여기서 Exponential(θ_1,θ_2)의 정의는

$$f(x;\theta_1,\theta_2) = \frac{1}{\theta_2}\exp\left\{-\frac{x-\theta_1}{\theta_2}\right\}, \quad x \ge \theta_1$$

입니다. 그러면 가능도 $L(\theta_1,\theta_2)$가

$$\begin{aligned} L(\theta_1,\theta_2; x_1,\cdots,x_n) &= \prod_{i=1}^n \frac{1}{\theta_2}\exp\left\{-\frac{x_i-\theta_1}{\theta_2}\right\} \\ &= \frac{1}{\theta_2^n}\exp\left\{-\frac{\sum_{i=1}^n (x_i-\theta_1)}{\theta_2}\right\}, \ \theta_1 \le x_1, \cdots, \theta_1 \le x_n \end{aligned}$$

으로 표현됩니다. 따라서

$$t_1 = \min(x_1, \cdots, x_n), \ t_2 = \sum_{i=1}^n (x_i - t_1)$$

으로 잡으면

$$L(\theta_1,\theta_2\,;x_1,\cdots,x_n) = \frac{1}{\theta_2^n}\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-t_1)}{\theta_2} - \frac{n\,(t_1-\theta_1)}{\theta_2}\right\}$$

$$= \frac{1}{\theta_2^n}\exp\left\{-\frac{t_2}{\theta_2} - \frac{n\,(t_1-\theta_1)}{\theta_2}\right\},\ \theta_1 \le t_1$$

가 됩니다. 그러므로 (T_1, T_2)가 (θ_1,θ_2)를 위한 결합충분통계량입니다. 이 때, T_1은 Exponential$(\theta_1,\theta_2/n)$ 분포를, T_2는 Gamma$(n-1,\theta_2)$ 분포를 따릅니다 (연습문제 4.10).

마지막 사례는 표본자료 $x_1,\cdots,x_n$에 대하여 Cauchy(θ) 분포를 가정한 경우입니다 (여기서 파라미터 θ는 분포중앙을 나타냅니다). 이 때 가능도는

$$L(\theta\,;x_1,\cdots,x_n) = \prod_{i=1}^{n}\left\{\frac{1}{\pi}\cdot\frac{1}{1+(x_i-\theta)^2}\right\}$$

입니다만, 통계량 T를 어떻게 잡아도 (9)의 형태로 인수화되지 않습니다. 그러므로 이 경우에서는 차원을 줄이는 자료축소가 가능하지 않습니다. 단, 이 경우에도 (연속형 확률변수의 경우에는 항상 그렇듯이) 순서통계량 $(X_{(1)},\cdots,X_{(n)})$은 결합충분합니다 (연습문제 4.11).

충분통계량이 존재하는 핵심적인 분포는 소위 지수족의 경우입니다. 지수족은 통계적 모형화와 추론에서 매우 중요한 역할을 하므로 잘 알아놓아야 할 것입니다.

지수족(指數族, exponential family) : 정의

확률변수 X의 확률(밀도)가

$$f(x\,;\theta_1,\cdots,\theta_k) = \exp\left\{\sum_{j=1}^{l}c_j(\theta_1,\cdots,\theta_k)\,P_j(x) + d(\theta_1,\cdots,\theta_k) + Q(x)\right\},$$

for $x\in A$ (여기서 함수의 정의역 A는 $(\theta_1,\cdots,\theta_k)$와 무관)

인 분포군(family of distributions)을 지수족의 확률분포라고 합니다. 통상적으로 k와 l은 일치합니다. ■

지수족의 예는 무척 많습니다.

1) 베르누이 Bernoulli(θ) 분포 :

$$f(x\,;\theta) = \theta^x(1-\theta)^{1-x} = \exp\{x\log_e\theta + (1-x)\log_e(1-\theta)\}$$

$$= \exp\left\{\log_e\frac{\theta}{1-\theta}\cdot x + \log_e(1-\theta)\right\},\quad x = 0, 1$$

이므로 베르누이 분포는

$$c(\theta) = \log_e \frac{\theta}{1-\theta},\ P(x) = x,\ d(\theta) = \log_e(1-\theta),\ A = \{0,1\}$$

인 지수족입니다.

2) 포아송 Poisson(θ) 분포 :

$$f(x\,;\theta) = \theta^x e^{-\theta}/x! = \exp\{\log_e \theta \cdot x - \theta - \log_e x!\},\ x = 0,1,2,\cdots$$

이므로 포아송 분포는

$$c(\theta) = \log_e \theta,\ P(x) = x,\ d(\theta) = -\theta,\ Q(x) = -\log_e x!,$$

$$A = \{0,1,2,\cdots\}$$

인 지수족입니다.

3) 정규분포 N(θ_1, θ_2) :

$$f(x\,;\theta_1,\theta_2) = \frac{1}{\sqrt{2\pi\,\theta_2}} \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2}\right\}$$

$$= \exp\left\{\frac{\theta_1}{\theta_2} \cdot x - \frac{1}{2\,\theta_2} \cdot x^2 - \frac{\theta_1^2}{2\,\theta_2} - \frac{1}{2}\log_e \theta_2 - \frac{1}{2} log_e(2\pi)\right\}$$

이므로

$$c_1(\theta_1,\theta_2) = \frac{\theta_1}{\theta_2}, \qquad P_1(x) = x\ ,$$

$$c_2(\theta_1,\theta_2) = -\frac{1}{2\,\theta_2},\ P_2(x) = x^2$$

인 지수족입니다.

4) 감마분포 Gamma(θ_1, θ_2) :

$$f(x\,;\theta_1,\theta_2) = \frac{1}{\Gamma(\theta_1)\,\theta_2^{\theta_1}}\, x^{\theta_1 - 1} \exp\left(-\frac{x}{\theta_2}\right)$$

$$= \exp\left[-\frac{1}{\theta_2} \cdot x + (\theta_1 - 1)\log_e x - \log_e\left\{\Gamma(\theta_1)\,\theta_2^{\theta_1}\right\}\right],\quad x \ge 0$$

이므로 감마분포는

$$c_1(\theta_1,\theta_2) = -\frac{1}{\theta_2}, \qquad P_1(x) = x\ ,$$

$$c_2(\theta_1,\theta_2) = \theta_1, \qquad P_2(x) = \log_e x\ ,$$

$$A = \{x \mid x \ge 0\}$$

인 지수족입니다.

물론 지수족이 아닌 분포의 예도 많습니다. 다음 분포들은 지수족이 아닙니다. 쉽게 확인할 수 있을 것입니다.

1) Uniform$(0,\theta)$ 분포.
2) Cauchy(θ) 분포.
3) Exponential(θ_1,θ_2) 분포.
4) Weibull(θ_1,θ_2) 분포.

지수족과 충분통계량 (exponential family and sufficient statistics) : 정리

확률변수 X가 지수족의 확률분포를 따르면

$$\left(\sum_{i=1}^{n} P_1(X_i), \cdots, \sum_{i=1}^{n} P_l(X_i)\right)$$

은 파라미터 $(\theta_1, \cdots, \theta_k)$를 위한 결합충분통계량(jointly sufficient statistics)입니다. 증명은 확장 인수화 정리에 의하여 명백합니다. ■

이 정리에 따라 지수족의 각 분포에서 파라미터 θ 또는 (θ_1,θ_2)에 대한 충분통계량은 다음과 같습니다.

1) 베르누이 Bernoulli(θ) 분포 : $T = \sum_{i=1}^{n} X_i$.

2) 포아송 Poisson(θ) 분포 : $T = \sum_{i=1}^{n} X_i$.

3) 정규분포 N(θ_1,θ_2) : $T_1 = \sum_{i=1}^{n} X_i$, $T_2 = \sum_{i=1}^{n} X_i^2$

앞에서는 $S_1 = \overline{X}$, $S_2 = \sum_{i=1}^{n} (X_i - \overline{X})^2$가 결합충분통계량이라고 하였었습니다. 어떻게 된 것일까요? 어렵지 않게 두 결합통계량이 1 : 1 대응함을 알 수 있습니다. 즉, (t_1,t_2)로부터 (s_1,s_2)를 정할 수 있고, 역으로도 할 수 있습니다.

4) 감마분포 Gamma(θ_1,θ_2) : $T_1 = \sum_{i=1}^{n} X_i$, $T_2 = \sum_{i=1}^{n} \log_e X_i$.

T_1과 T_2의 결합확률분포를 구하는 것은 탐구문제로 남깁니다.

4.A 연습문제

4.1 $x_1, \cdots, x_n$ 을 지수분포 Exponential(θ) 로 모형화한다고 합시다. 파라미터 θ 에 대한 최대가능도추정량을 구하시오. 또한 관측자료가

$$0.56, 0.06, 1.16, 2.16, 3.78, 1.90, 1.18, 1.06, 0.04, 2.98 \quad (n = 10)$$

일 때 정규화 가능도 $L(\theta)$ 를 그려보세요.

4.2 $x_1, \cdots, x_n$ 을 정규분포 Normal(θ, σ^2) 로 모형화한다고 합시다 (이 때 분산 σ^2 이 기지(旣知, known)임을 전제). 파라미터 θ 에 대한 최대가능도추정량을 구하시오. 또한 $\sigma^2 = 10$ 이고 관측자료가

$$59, 59, 66, 72, 45, 35, 45, 52, 48, 63 \quad (n = 10)$$

인 경우 정규화 가능도 $L(\theta)$ 를 그려보세요.

4.3 4.1절의 베르누이 자료 $x_1, \cdots, x_{100}$ (성공수 $s = 57$) 에 대하여 1/7 규칙에 의한 성공률 θ 의 구간이 $(0.471, 0.665)$ 임을 보이세요.

4.4 4.2절의 이산화 지수 표본자료 ($x = 1,1,1,1,1,1,2,3,4,6$; $n = 10$)에 대한 mle가 $\hat{\lambda} = 0.6466$ 이며 1/7 규칙에 의한 λ 의 구간이 $(0.320, 1.15)$ 임을 보이세요.

4.5 $x_1, \cdots, x_n$ 을 균일분포 Uniform($\theta-\alpha, \theta+\alpha$) 로 모형화한다고 합시다 (이 때, α 는 기지(旣知, known)). 파라미터 θ 에 대한 최대가능도추정량을 구하시오.

4.6 Hardy-Weinberg Law에 따르면 혈액형 M, MN, N에 대한 확률이 각각

$$(1-\theta)^2, \ 2(1-\theta)\theta, \ \theta^2$$

이여야 한다. 실제 n 명의 관측한 결과 혈액형 M, MN, N의 빈도가 N_1, N_2, N_3 였습니다 ($N_1 + N_2 + N_3 = n$). θ 에 대한 최대가능도추정량을 구하시오.

4.7* $x_1, \cdots, x_n$ 을 감마분포 Gamma(θ_1, θ_2) 로 모형화할 때 (θ_1, θ_2) 를 구하기 위한 방정식을 유도하십시오. 자료가

$$0.6, 8.5, 8.6, 9.5, 3.4, 4.0, 20.0, 13.8, 14.8, 11.9 \quad (n = 10)$$

로 관측되었을 때, (θ_1, θ_2) 에 대한 mle를 수치적으로 구하세요.

4.8* (앞 문제의 계속) 감마분포 표본자료의 가능도 $L(\theta_1, \theta_2)$ 와 1/20 규칙에 의한 (θ_1, θ_2) 영역을 그래프로 표현해보세요.

[수학 소프트웨어인 Mathematica를 사용하는 경우엔 Plot3D, FindMinimum, ContourPlot 등의 함수를 쓰십시오].

4.9 $X_1, \cdots, X_n$이 파레토 분포 Pareto(θ)로부터의 iid 확률변수라고 합시다. 즉, 이들 변수 각각의 확률밀도함수는

$$f_X(x\,;\theta) = \frac{\theta}{x^{\theta+1}}, \quad \theta > 1,\ x \geq 1$$

입니다. θ를 위한 충분통계량을 찾고 그것의 분포를 말해 보세요.

답. $\prod_{i=1}^{n} X_i$ 또는 $\sum_{i=1}^{n} \log_e X_i \ (= T)$, $T \sim$ Gamma$(n, 1/\theta)$.

4.10 $X_1, \cdots, X_n$이 Exponential(θ_1, θ_2)분포로부터의 iid 확률변수라고 합시다. 이 때

$$T_1 = \min(X_1, \cdots, X_n),\ T_2 = \sum_{i=1}^{n} (X_i - T_1)$$

이 독립적으로 각각 Exponential$(\theta_1, \theta_2/n)$분포와 Gamma$(n-1, \theta_2)$분포를 따름을 보이세요.

4.11 $X_1, \cdots, X_n$이 연속형 확률분포 $F(.)$로부터의 iid 확률변수라고 할 때, 순서통계량 $X_{(1)} \leq \cdots \leq X_{(n)}$이 언제나 결합충분함을 보이세요.

4.12 $X_1, \cdots, X_n$이 와이블 분포 Weibull(θ_1, θ_2)로부터의 iid 확률변수라고 합시다. (θ_1, θ_2)를 위한 결합충분통계량을 구하세요.

<u>탐구문제</u> $X_1, \cdots, X_n$이 감마분포 Gamma(θ_1, θ_2)로부터의 iid 확률변수라고 합시다. (θ_1, θ_2)를 위한 결합충분통계량인 $T_1 = \sum_{i=1}^{n} X_i$와 $T_2 = \sum_{i=1}^{n} \log_e X_i$의 결합확률분포를 구하세요.

4.B 읽을만한 책

수치계산에 보충 참고문헌이 필요하면 다음 책을 보기 바랍니다.

- 손건태 (1996)「전산통계개론」 자유 아카데미. (9장)
- 최영훈 · 이승천 (1995)「C에 의한 전산통계」 자유 아카데미. (3장)
- Burden, R.L. and Faires, J.D. (1997) *Numerical Analysis*, 6th Edition. Brooks and Cole. (Chapter 2)

가능도(likelihood)와 충분통계량에 대하여는 다음 책을 보십시오.

- Hogg, R.V. and Craig, A.T. (1995) *Introduction to Mathematical Statistics*, 5th Edition. Prentice Hall. (Chapter 7)
- Casella, G. and Berger, R.L. (1990) *Statistical Inference*. Duxbury. (Chapter 6)
- Cox, D.R. and Hinkley, D.V. (1974) *Theoretical Statistics*. Chapman and Hall. (Chapter 2)

이 장의 <그림 6>와 <그림 7>은 수학계산 소프트웨어인 Mathematica를 써서 얻은 것입니다. 다른 소프트웨어로도 될 것 같군요. 시도해보세요.