5장 모수적 모델과 지수 패밀리

2019.08.14

신 현석

들어가며

전통적 통계학 방법론 : 저차원 모수, 확률밀도 → 수학적 용이

$$\mathcal{F} = \{f_\mu(x); x \in \mathcal{X}, \mu \in \Omega\}$$

확률밀도함수

모수 20이하

가장 보편적 모수적 모델

지수분포 : 전통 - 컴퓨터 시대의 연결점

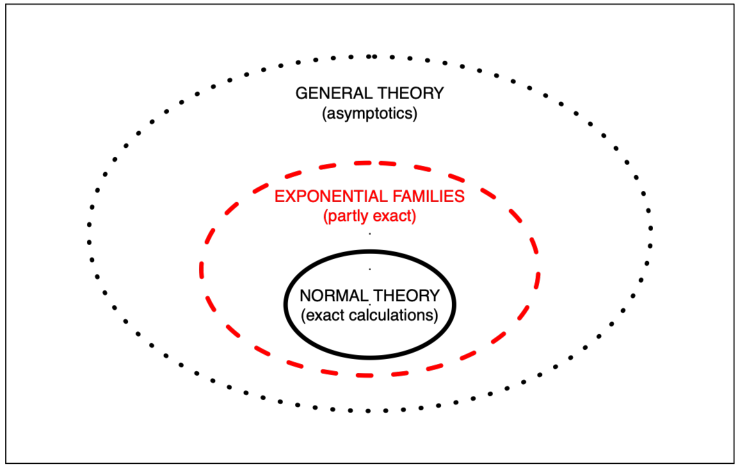


Figure 8.3 Three levels of statistical modeling.

5.1 일변량 패밀리

정규, 포아송, 이항, 감마, 베타

Name, Notation	Density	$\mathcal{X}$	Ω	Expectation, Variance
Normal $\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	$\mathcal{R}^1$	$\mu \in \mathcal{R}^1$ $\sigma^2 > 0$	μ σ^2
Poisson $\mathrm{Poi}(\mu)$	$\frac{e^{-\mu}\mu^x}{x!}$	$\{0, 1, \ldots\}$	$\mu > 0$	μ μ
Binomial $\mathrm{Bi}(n, \pi)$	$\frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}$	$\{0, 1, \ldots, n\}$	$0 < \pi < 1$	$n\pi$ $n\pi(1-\pi)$
Gamma $\mathrm{Gam}(\nu, \sigma)$	$\frac{x^{\nu-1}e^{-x/\sigma}}{\sigma^\nu\Gamma(\nu)}$	$x \geq 0$	$\nu > 0$ $\sigma > 0$	$\sigma\nu$ $\sigma^2\nu$
Beta $\mathrm{Be}(\nu_1, \nu_2)$	$\frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} x^{\nu_1-1}(1-x)^{\nu_2-1}$	$0 \leq x \leq 1$	$\nu_1 > 0$ $\nu_2 > 0$	$\nu_1/(\nu_1+\nu_2)$ $\frac{\nu_1\nu_2}{(\nu_1+\nu_2)^2(\nu_1+\nu_2+1)}$

지수

카이제곱

이항분포: 성공확률이 p, 실패확률이 $q=1-p$인 베르누이 시행의 n회의 독립시행에서 성공횟수를 나타내는 이항확률변수 X의 확률분포는

$b(x; n, p) = \binom{n}{x} p^x q^{n-x}$, $x = 0, 1, 2, \cdots, n$ 으로 주어진다.

포아송 분포 : 주어진 시간간격 t 동안에, 혹은 일정 영역 t에서 발생하는 결과의 수를 나타내는

포아송 확률변수 X의 확률분포 : $p(x; \lambda t) = \dfrac{e^{\lambda t}(\lambda t)^x}{x!}$, $x = 0, 1, 2, \cdots$

여기서 λ는 단위시간 또는 단위면적에서 발생하는 결과의 수, $e = 2.71828\cdots$

*** 이항분포와 포아송 분포

이항분포에서, $n \to \infty$, $p \to 0$, np는 상수인 경우, 포아송 분포에 가까워짐.
n이 크고 p가 0에 가까운 값이면 이항분포를 포아송 분포로 근사. (p가 1에 가까우면, 성공과 실패를 바꾸어 근사)

*** 이항분포와 포아송 분포

이항분포에서, $n \to \infty$, $p \to 0$, np는 상수인 경우, 포아송 분포에 가까워짐.
n이 크고 p가 0에 가까운 값이면 이항분포를 포아송 분포로 근사. (p가 1에 가까우면, 성공과 실패를 바꾸어 근사)

증명) $p = \dfrac{\mu}{n}$이므로

$$
\begin{aligned}
b(x; n, p) &= \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \frac{n(n-1)\cdots(n-x+1)}{x!} p^x (1-p)^{n-x} \\
&= \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\mu}{n}\right)^x \left(1-\frac{\mu}{n}\right)^{n-x} \\
&= 1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\frac{\mu^x}{x!}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}
\end{aligned}
$$

$n \to \infty$이면 (x, μ는 상수)

$$\lim_{n\to\infty} 1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x-1}{n}\right) = 1,\ \lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{-x} = 1,$$

$$\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n = \lim_{n\to\infty}\left(1+\frac{-\mu}{n}\right)^n = e^{-\mu}$$

($\because$ $e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$이므로 $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = \lim_{n\to\infty}\left(\left(1+\frac{1}{n/x}\right)^{\frac{n}{x}}\right)^x = e^x$)

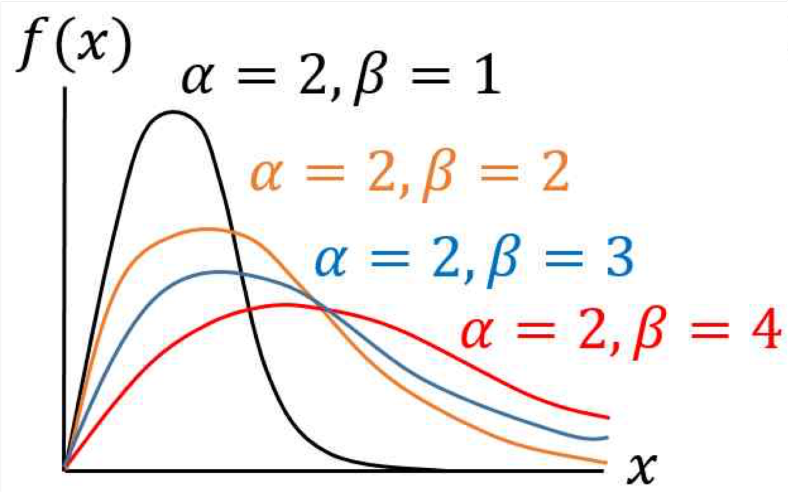
그러므로 $b(x; n, p) \to 1 \cdot \dfrac{\mu^x}{x!} e^{-\mu} \cdot 1 = \dfrac{e^{-\mu}\mu^x}{x!} = p(x; \mu)$이다.

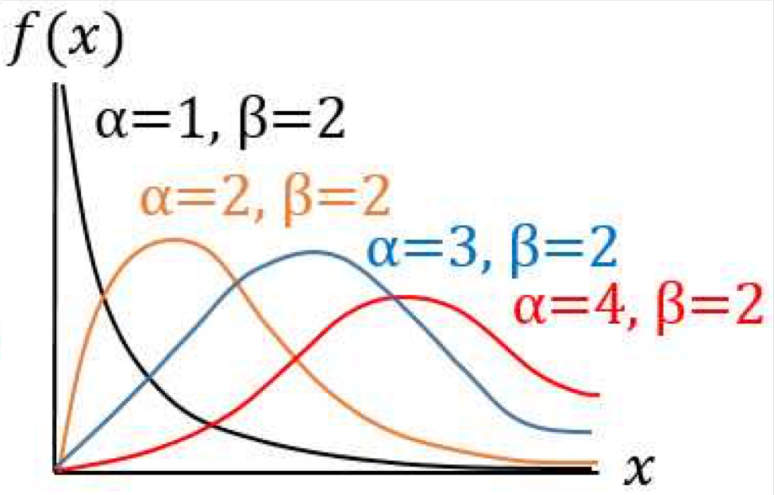
감마분포 : 연속확률변수 X의 확률밀도함수가

$$f(x;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{\beta}}, & x>0\\ 0 & \text{그 외의 } x\end{cases}$$

$(\alpha>0,\ \beta>0)$

와 같이 주어질 때 X는 모수 α, β 를 가지는 감마분포를 따른다고 한다.





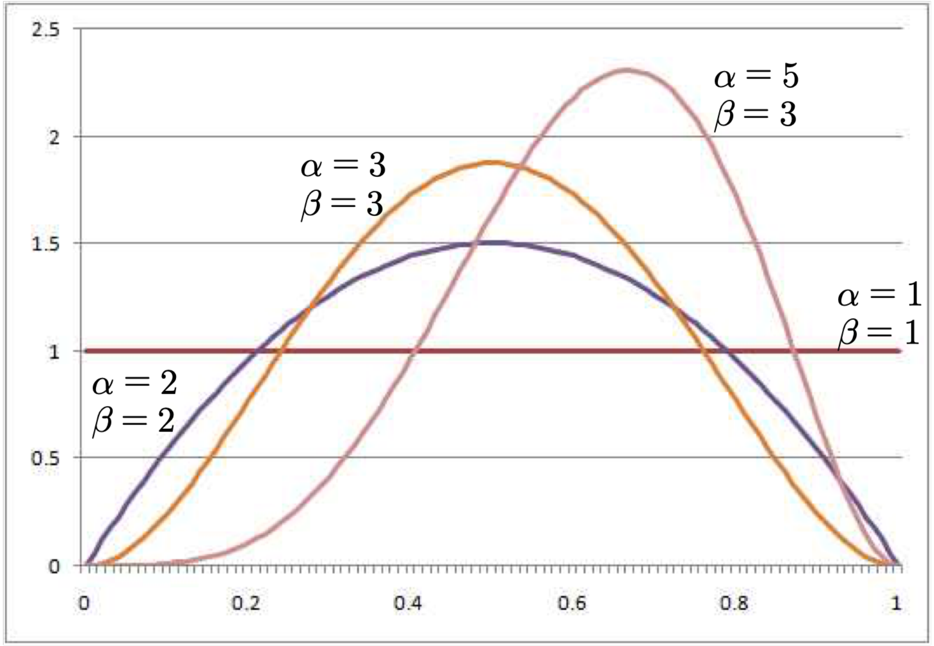
(α : 모양을 결정 - shape parameter, β : 크기를 결정 - scale parameter)

* $\alpha=1$인 특수한 감마분포를 지수분포라 한다.
  카이제곱 분포(chi-square distribution) : $\alpha=\dfrac{v}{2}$(v는 양의 정수), $\beta=2$인 경우의 감마분포

베타분포 : 확률밀도함수가

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

인 분포를 모수가 α, β인 베타분포라 한다.



5.2 다변량 정규분포

여러 변수 - 하나의 벡터
예) 신체검사 - 나이, 키, 몸무게, 혈압
예) 입시 - 시험성적, 졸업성적, 면접성적

- 각 벡터의 원소들 사이에 연관성 존재.

$$x \sim \mathcal{N}_p(\mu, \mathbf{\Sigma})$$

$$\mu = E\{x\} = \left(E\{x_1\}, E\{x_2\}, \ldots, E\{x_p\}\right)'$$

$$\begin{aligned}\mathbf{\Sigma} &= E\left\{(x-\mu)(x-\mu)'\right\} = \left(E\left\{(x_i-\mu_i)(x_j-\mu_j)\right\}\right)\\ &= \boldsymbol{T}\boldsymbol{T}'\end{aligned}$$

* X의 평균벡터: $\mu = (\mu_1, \mu_2, \cdots, \mu_p)^t = \begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_p\end{bmatrix}$,

X의 공분산행렬(분산행렬): $\Sigma = \begin{bmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p}\\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp}\end{bmatrix} = \left[\,Cov(X_i, X_j)\,\right]$

$$f_{\mu,\boldsymbol{\Sigma}}(x) = \frac{(2\pi)^{-p/2}}{|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\boldsymbol{\Sigma}^{-1}(x-\mu)}$$

상수 : (1 x p) x (p x p) x (px1) = 1 x 1

$$f(z) = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}\sum_1^p z_i^2} = (2\pi)^{-\frac{p}{2}} e^{-\frac{1}{2}z'z}$$

$$x = \mu + \boldsymbol{T} z$$

[p. 57] *Formula* (5.12). From $z = \boldsymbol{T}^{-1}(x - \mu)$ we have $dz/dx = \boldsymbol{T}^{-1}$ and

$$f_{\mu,\boldsymbol{\Sigma}}(x) = f(z)|\boldsymbol{T}^{-1}| = (2\pi)^{-\frac{p}{2}}|\boldsymbol{T}^{-1}|e^{-\frac{1}{2}(x-\mu)'\boldsymbol{T}^{-1\prime}\boldsymbol{T}^{-1}(x-\mu)}, \quad (5.63)$$

so (5.12) follows from $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{\Sigma}$ and $|\boldsymbol{T}| = |\boldsymbol{\Sigma}|^{1/2}$.

If $p_1 = p_2 = 1$, then (5.18) reduces to

$$x_2 | x_1 \sim \mathcal{N}\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1),\ \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)$$

* $X = x$일 때 Y의 조건부 평균이 x에 대해 선형인 경우($E(Y|x) = a + bx$):

이산형인 경우 $a + bx = E(Y|x) = \sum_y yf(y|x) = \sum_y y\frac{f(x,y)}{f_X(x)}$ 이므로,

$$\sum_y yf(x,y) = (a+bx)f_X(x)$$

$$\sum_x \sum_y yf(x,y) = \sum_x (a+bx)f_X(x)$$

$$\mu_Y = a + b\mu_X \qquad (1)$$

또한, $\sum_y yf(x,y) = (a+bx)f_X(x)$ 에서

$$\sum_x \sum_y xyf(x,y) = \sum_x (ax+bx^2)f_X(x)$$

$$E(XY) = aE(X) + bE(X^2)$$

* $E(XY) = \sigma_{XY} + \mu_X\mu_Y$, $\sigma_{XY} = \rho\sigma_X\sigma_Y$, $E(X^2) - \mu_X^2 = \sigma_X^2$ 이므로

$$\mu_X\mu_Y + \rho\sigma_X\sigma_Y = a\mu_X + b(\mu_X^2 + \sigma_X^2) \qquad (2)$$

(1)과 (2):

$$\mu_Y = a + b\mu_X$$

$$\mu_X\mu_Y + \rho\sigma_X\sigma_Y = a\mu_X + b(\mu_X^2 + \sigma_X^2)$$

연립방정식의 해는 $a = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X$, $b = \rho\frac{\sigma_Y}{\sigma_X}$

그러므로 $E(Y|x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$

If $p_1 = p_2 = 1$, then (5.18) reduces to

$$x_2|x_1 \sim \mathcal{N}\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1),\ \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)$$

$$\begin{aligned}
\sigma^2_{Y|x} &= \int_{-\infty}^{\infty} (y - E(Y \mid x))^2 f(y \mid x) dy \\
&= \int_{-\infty}^{\infty} \left(y - \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right)^2 f(y \mid x) dy
\end{aligned}$$

$$\int_{-\infty}^{\infty} \sigma^2_{Y|x} f_X(x) dx = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} \left(y - \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right)^2 f(y \mid x) dy\right) f_X(x) dx$$

$$\begin{aligned}
\sigma^2_{Y|x} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(y - \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right)^2 f(y \mid x) f_X(x) dy dx \qquad \text{(분산은 상수)} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(y - \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right)^2 f(x,\ y) dy dx \\
&= E\left((Y - \mu_Y)^2 - 2\rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)(Y - \mu_Y) + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}(X - \mu_X)^2\right) \\
&= \sigma_Y^2 - 2\rho\frac{\sigma_Y}{\sigma_X}\sigma_{XY} + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_X^2 \\
&= \sigma_Y^2 - 2\rho\frac{\sigma_Y}{\sigma_X}\rho\sigma_X\sigma_Y + \rho^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_X^2 \\
&= \sigma_Y^2 - 2\rho^2\sigma_Y^2 + \rho^2\sigma_Y^2 = \sigma_Y^2(1 - \rho^2)
\end{aligned}$$

* $p=2$인 경우 이변량 정규분포

$$\Sigma=\begin{bmatrix}\sigma_{11}\ \sigma_{12}\\ \sigma_{21}\ \sigma_{22}\end{bmatrix},\ |\Sigma|=\sigma_{11}\sigma_{22}-\sigma_{12}^2,\ \Sigma^{-1}=\frac{1}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}\begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{bmatrix},\ \rho_{12}=\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}},$$

$\sigma_{12}=\rho_{12}\sqrt{\sigma_{11}\sigma_{22}}$ 이므로,

$|\Sigma|=\sigma_{11}\sigma_{22}-\sigma_{12}^2=\sigma_{11}\sigma_{22}-\rho_{12}^2\sigma_{11}\sigma_{22}=\sigma_{11}\sigma_{22}(1-\rho_{12}^2)=\sigma_1^2\sigma_2^2(1-\rho^2)$이고,

$$(x-\mu)^t\Sigma^{-1}(x-\mu)=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\left[x_1-\mu_1\ x_2-\mu_2\right]\begin{bmatrix}\sigma_{22} & -\sigma_{12}\\ -\sigma_{12} & \sigma_{11}\end{bmatrix}\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix}$$

$$=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\left[(x_1-\mu_1)\sigma_{22}-(x_2-\mu_2)\sigma_{12}\quad -(x_1-\mu_1)\sigma_{12}+(x_2-\mu_2)\sigma_{11}\right]\begin{bmatrix}x_1-\mu_1\\ x_2-\mu_2\end{bmatrix}$$

$$=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\left[(x_1-\mu_1)^2\sigma_{22}-(x_1-\mu_1)(x_2-\mu_2)\sigma_{12}-(x_1-\mu_1)(x_2-\mu_2)\sigma_{12}+(x_2-\mu_2)^2\sigma_{11}\right]$$

$$=\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\left((x_1-\mu_1)^2\sigma_2^2-2(x_1-\mu_1)(x_2-\mu_2)\sigma_{12}+(x_2-\mu_2)^2\sigma_1^2\right)$$

$$=\frac{1}{(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\sigma_1\sigma_2\frac{1}{\sigma_1^2\sigma_2^2}(x_1-\mu_1)(x_2-\mu_2)+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)$$

$$=\frac{1}{(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)$$

그러므로 $f(x_1,x_2)=\dfrac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)}e^{-\frac{Q}{2}}$,

$$Q=\frac{1}{(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)$$

5.3 다모수 패밀리의 피셔 정보 경계

5.4 다항분포 $\boldsymbol{x} \sim \mathrm{Mult}_L(n, \boldsymbol{\pi})$

$$f_{\boldsymbol{\pi}}(\boldsymbol{x}) = \frac{n!}{x_1! x_2! \dots x_L!} \prod_{l=1}^{L} \pi_l^{x_l}$$

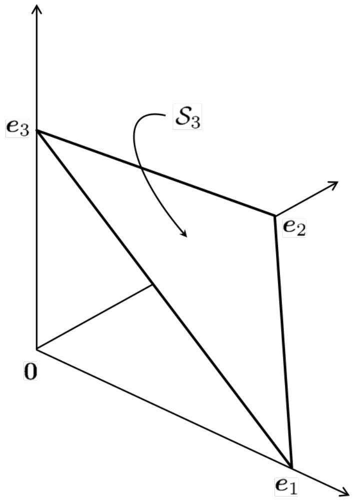
관측치가 오직 유한한 개수의 이산 값

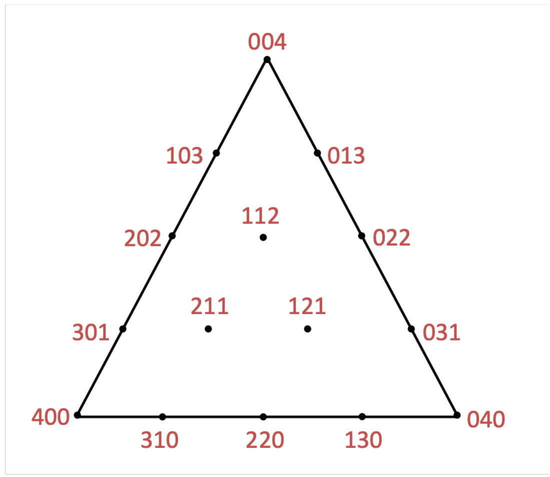
	success	failure
new	1 **9**	2 **12**
old	3 **7**	4 **17**

$L = 4$

$$\boldsymbol{x} = (9, 12, 7, 17)'$$

$$x_l = \#\{\text{cases having outcome } l\}$$





$$\boldsymbol{e}_l = (0, 0, \dots, 0, 1, 0, \dots, 0)'$$

이항분포: 성공확률이 p, 실패확률이 $q=1-p$인 베르누이 시행의 n회의 독립시행에서 성공횟수를 나타내는 이항확률변수 X의 확률분포는

$b(x;n,p)=\binom{n}{x}p^{x}q^{n-x},\ x=0,1,2,\cdots,n$ 으로 주어진다.

다항분포 : 각 시행에서 $p_1, p_2, \cdots, p_k$의 확률로 $E_1, E_2, \cdots, E_k$ 중 어느 하나가 발생할 경우, n번의 독립시행에서 각각 $E_1, E_2, \cdots, E_k$의 발생횟수를 나타내는 확률변수 $X_1, X_2, \cdots, X_k$의 확률분포는 $f(x_1, x_2, \cdots, x_k; p_1, p_2, \cdots, p_k, n)=\binom{n}{x_1, x_2, \cdots, x_k}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$

$\left(\sum_{i=1}^{k}x_i=n,\ \sum_{i=1}^{k}p_i=1\right)$

5.5 지수 패밀리

$$\frac{f_\mu(x)}{f_{\mu_0}(x)} = e^{-(\mu-\mu_0)}\left(\frac{\mu}{\mu_0}\right)^x \longrightarrow f_\mu(x) = e^{\alpha x-\psi(\alpha)} f_{\mu_0}(x)$$

$$\alpha = \log\{\mu/\mu_0\}$$

$$\psi(\alpha) = \mu_0(e^\alpha - 1)$$

$$\tilde{f}_\mu(x) = e^{\alpha x} f_{\mu_0}(x) \qquad e^{\psi(\alpha)} = \sum_0^\infty e^{\alpha x} f_{\mu_0}(x)$$

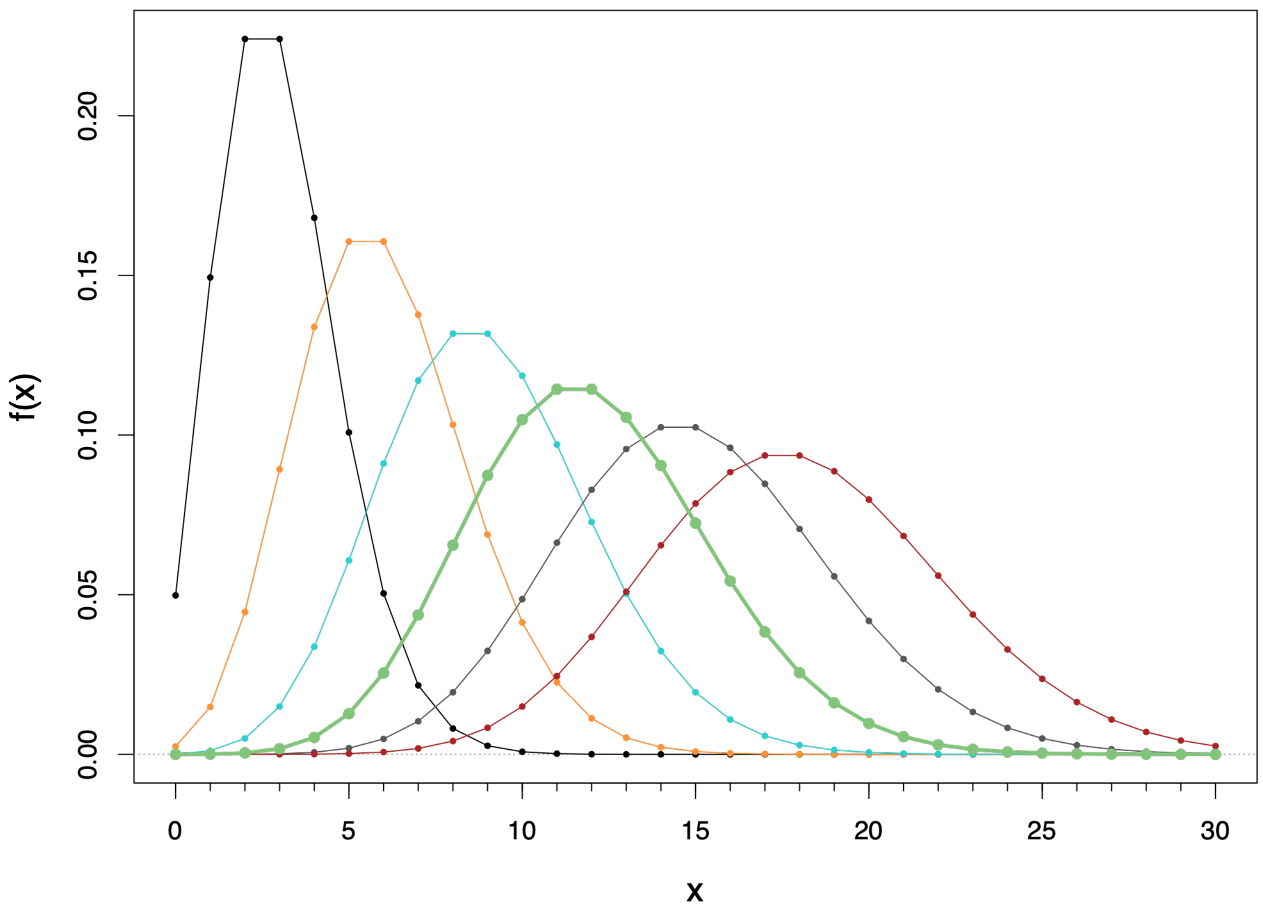


Figure 5.6 Poisson densities for $\mu = 3, 6, 9, 12, 15, 18$; heavy green curve with dots for $\mu = 12$.

$$
\begin{aligned}
f_\alpha(\boldsymbol{x}) &= \prod_{i=1}^{n} e^{\alpha' y_i - \psi(\alpha)} f_0(x_i) \\
&= e^{n(\alpha' \bar{y} - \psi(\alpha))} f_0(\boldsymbol{x}),
\end{aligned}
$$

왜 다른 변환 형식을 고려하지 않고 지수적 기울임에 관심을 두는 것일까? 그 대답은 반복된 표본추출과 관련 있다.

n이 얼마나 커지든 모든 추론 정보를 p차원 통계량 y(bar)에 압축할 수 있다. → 수학적 용이성