Lecture Note 3.

Conditional Independence.

$X_1 \perp X_2 \mid Y \;\not\Leftrightarrow\; X_1 \perp X_2$.

Bayesian Inference.

$$f(\theta|x) = \frac{f(x|\theta)\cdot f(\theta)}{f(x)} = \frac{f(x|\theta)\cdot f(\theta)}{\int f(x|\theta)\cdot f(\theta)\,d\theta} \propto \underset{\text{likelihood}}{f(x|\theta)}\cdot \underset{\text{prior}}{f(\theta)}$$

post.

Sufficiency.

$X_1 \cdots X_n \mid S \perp \theta$. S is Sufficient Statistic for θ.

쉽게말해, $S(\underline{X})$가 θ의 정보를 모두 가지고 있음.

예를들어서 μ, σ^2 에 의해 모든 분포가 결정되는 정규분포의 경우,

충분통계량은 $\hat{\mu} = \sum X_i$ $\hat{\sigma}^2 = \sum X_i^2$ jointly.

Factorization Theorem!

let $p(x|\theta)$ be joint pdf/pmf of sample x.

$S(x)$ be sufficient statistic for θ

when $\exists\, g(S(x)|\theta), h(x)$. $\forall x$. $\forall \theta \in \Theta$.

$$p(x|\theta) = \underbrace{g(S(x)|\theta)}_{\theta\text{에 의존하는 부분}} \cdot \underbrace{h(x)}_{x\text{로만 이루어진 부분}}$$

주로 식분해를 통해 충분통계량 $S(x)$를 찾는데 쓰임

사후분포는 오로지 충분통계량을 통해 데이터 X에 의존한다.

$$p(\theta|X) \propto p(x|\theta)\cdot p(\theta) = g(S(x)|\theta)\cdot h(x)\cdot p(\theta) \propto g(S(x)|\theta)\cdot p(\theta).$$

$$p(\theta|X) = p(\theta|S) \quad (S(x) = s).$$

Beta Dist.

Suppose $X_1 \cdots X_n \mid \theta \overset{iid}{\sim} \mathrm{Bernoulli}(\theta)$.

$$f_X(x) = \prod_{i=1}^{n} f_X(x_i) = \prod \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}.$$

$$= \exp\left[\underline{\sum x_i}\;\underline{\ln\theta} + (n-\sum x_i)\ln(1-\theta)\right]\cdot \underline{\prod_{i=1}^{n} I_{\{0,1\}}(x_i)}.$$

$$= g(S(x)|\theta)\cdot \underline{h(x_1, \cdots x_n)}$$

θ에 의존하는 $S(x)$는 $\sum x_i$ $\therefore S(x) = \sum x_i$ is SS for θ.

원하는 posterior를 얻기위해 prior를 어떻게 설정할거냐?

베르누이 시행을 위해 Commonly used 되는 사전분포가 Beta분포임.

베타분포란,

$\cancel{\theta} \sim Beta(\alpha, \beta)$.

$$P(\cancel{\theta}_i \mid \alpha,\beta) = \frac{1}{B(\alpha,\beta)} \cdot x_i^{\alpha-1} \cdot (1-x_i)^{\beta-1} \cdot I_{(0,1)}(x_i). \qquad Beta(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

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

$$E(X\mid\alpha,\beta) = \int_0^1 x\cdot p(x\mid\alpha,\beta) = \frac{\alpha}{\alpha+\beta}.$$

$$Var(X\mid\alpha,\beta) = \frac{\alpha}{\alpha+\beta}\,\frac{\beta}{(\alpha+\beta)(\alpha+\beta+1)}.$$

Mostly for R.V. X. (proportion).

From Beta prior to Beta posterior.

Beta family is used to express experimenter's uncertainty about X (proportion).

Suppose we chose Beta prior with params α,β. and $X_1 \cdots X_n \mid \theta \overset{iid}{\sim} Bernoulli(\theta)$.

Then, posterior should be

$$p(\theta\mid x_1\cdots x_n) \propto \underbrace{p(x_1\cdots x_n\mid\theta)}_{\substack{(\Pi\, Bernoulli(\theta))\\ \text{Likelihood}}} \cdot \underbrace{p(\theta)}_{\substack{(Beta)\\ \text{prior}}}.$$

$$= \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i} \times \theta^{\alpha-1}\cdot(1-\theta)^{\beta-1}.$$

$$= \theta^{\Sigma x_i+\alpha-1}\cdot(1-\theta)^{n-\Sigma x_i+\beta-1} \sim \overset{\text{pdf of}}{Beta}(\Sigma x_i+\alpha,\ n-\Sigma x_i+\beta).$$

$$E(\theta\mid X,\alpha,\beta) = \frac{\alpha+\Sigma x_i}{\alpha+\beta+n} = \frac{\alpha+\beta}{\alpha+\beta+n}\cdot\left(\frac{\alpha}{\alpha+\beta}\right) + \frac{n}{\alpha+\beta+n}\cdot\left(\frac{x}{n}\right) = \text{proportion of } \underbrace{\text{prior Dist.}}_{p(\theta)}\ E(\theta) + \text{proportion of } \underbrace{\text{Likelihood}}_{p(X\mid\theta)}\ E(X\mid\theta).$$

which was $Beta(\alpha,\beta)$. , $\Pi\, Bernoulli(\theta)$.

$$Var(\theta\mid X,\alpha,\beta) = \sim$$

결과적으로, 사전분포 : $Beta(\alpha,\beta)$.

사후분포 : $Beta(\alpha+\Sigma x_i,\ n-\Sigma x_i+\beta)$.

로써, 사전과 사후 분포가 동일한 분포형태를 갖게 되었다.

이는 prior를 Beta로 설정해준 행위에서 기인한다.

이렇게 사전/사후 분포가 동일한 형태인 것을 conjugate(켤레)하다고 한다.

코드로 구현하는 방법

grid approximation.

Monte Carlo using built in pdf (exact)

Bayesian Inference.

Lecture Note 4.

MAP : $\underset{\theta}{\operatorname{argmax}}\ p(\theta|X)$. $\equiv$ posterior Mode (최빈값).

Loss function : $L(\theta, a)$. $\theta \in \underset{\text{(sample space)}}{\Omega}$. $a \in \mathbb{R}$ (action).

$$E_{p(\theta)}[L(\theta,a)] = \int_{\Omega} L(\theta,a)\cdot p(\theta)\, d\theta \quad (\text{ex. } \theta: \text{proportion})$$

$$E_{p(\theta|X)}[L(\theta,a)|X] = \int_{\Omega} L(\theta,a)\cdot p(\theta|x)\, d\theta$$

Bayes Estimator.

$\boxed{\underset{\delta(x)}{\operatorname{argmin}}\ E[L(\theta,a)|X] = \delta^*(x).}$ is called "Bayes estimator of θ".

(Bayes Action).

Expected Loss principle. :

estimate that minimizes expected Loss is best.

Suppose $a_1 = \delta_1(X)$. $a_2 = \delta_2(X)$. We prefer action that posterior expected loss $E[L(\theta,a)|X]$ is smaller.

a is function of X.

① $L(\theta,a) = (\theta-a)^2$. $\delta^*(X) = \min E[(\theta-a)^2|X]$.

$\underset{a}{\operatorname{argmin}}\ E[L(\theta,a)|X] = \delta^*(X)$. $\quad E_{\theta}[(\theta-a)^2|X] = \int_{\Omega} (\theta-a)^2 \cdot p(\theta|x)\, d\theta$. $\quad \delta^*(x) = E(\theta|X)$.

Expectation of conditional expectation.

$$E(E(\theta|X)|X) = \int E(\theta|X=x)\cdot f_X(x)\, dx = \iint \theta f_{\theta|X=x}(\theta)\, d\theta \cdot f_X(x)\, dx$$

$$= \iint \theta \cdot f_{\theta,X}(\theta,x)\, dx\, d\theta. = \int \theta \cdot \underbrace{\int f_{\theta,X}(\theta,x)\, dx}_{\text{marginalize}}\, d\theta.$$

$$= \int \theta \cdot f_{\theta}(\theta)\cdot d\theta \quad = E(\theta).$$

② $L(\theta,a) = |\theta - a|$ (이런식으로 증명).

$\delta^*(X) = \min E(|\theta-a|\,|X)$

let $\varphi(a) = E_{\theta}(|\theta-a|) = \int_{\cancel{\theta\geq a}\ a}^{\infty} (\theta-a)\pi(\theta)\cdot d\theta + \int_{\cancel{\theta\leq a}\ -\infty}^{a} (a-\theta)\pi(\theta)\, d\theta$

$$\varphi'(a) = -\int_a^{\infty} \pi(\theta)\, d\theta + \int_{-\infty}^{a} \pi(\theta)\, d\theta.$$

$$= -P(\theta \geq a) + P(\theta \leq a) \overset{\text{set}}{=} 0,$$

$P(\theta \leq a) = P(\theta \geq a)$, a is Median of dist. θ (post.).

③ $L(\theta, a) = \lambda(\theta) \cdot (\theta - a)^2$. $\lambda(\theta)(\theta - a)$

(magnitude that depend on θ).

④ $L(\theta, a) = \begin{cases} C_1 |\theta - a| & \theta \geq a \\ C_2 |\theta - a| & \theta < a. \end{cases}$ $\quad C_1 P(\theta \geq a | X) = C_2 \cdot P(\theta \leq a | X)$

Credible Intervals.

C.I : $1-\alpha$ intervals contain true value of θ. (고정 true parameter is in the CI with $100(1-\alpha)\%$ ~~probability~~).

Credible C.I : $1-\alpha$ true value of θ is contained. in Credible C.I. (〃 ○)

$$P(\theta_L < \theta < \theta_U | X) = \int_{\theta_L}^{\theta_U} p(\theta | X) \cdot d\theta = 1 - \alpha.$$

Definition.

모수공간 Θ의 부분집합 C가 유의수준 α에 대해 $P(\theta \in C | X) = \int_C p(\theta | X) d\theta \geq 1 - \alpha$. 일때, C를 given X의 θ에 대한 Credible interval 이라고 한다.

WinBUGS

- Model specification
- data set
- initial values for MCMC.

Predictive Distributions.

Lecture Note 5.

$$p(\tilde{x}|x) = \int p(\tilde{x},\theta|x)\,d\theta = \int \underline{p(\tilde{x}|\theta,x)}\cdot p(\theta|x)\,d\theta$$

(marginalize).

$\tilde{x} \perp x | \theta.$

$$= \int p(\tilde{x}|\theta)\cdot p(\theta|x)\,d\theta.$$

$(X_1 \cdots X_n) = \underline{X}$: 동전던지기의 결과.

$X_i|\theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$, $\theta \sim \text{Unif}(0,1) \equiv \text{Beta}(1,1)$ prior: $p(\theta)$

$X = \sum_{\forall i} X_i \sim \text{Bin}(n,\theta)$. likelihood $p(x|\theta)$.

$\theta|x \sim \text{Beta}(x+1, n-x+1)$ posterior $p(\theta|x)$

$\tilde{x} \sim ?$ $\quad p(\tilde{x}) = \int_0^1 p(\tilde{x}|\theta)\,p(\theta)\,d\theta.$ prior predictive dist.

(marginalize). 일반적으로 $\tilde{x}|\theta$가 given.

$p(\tilde{x}|x) \sim ?$

Beta-Binomial.

$\theta \sim \text{Beta}(\alpha,\beta)$, $X|\theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$, X_{n+1}.

(generalized $\underline{X}$).

sol.

$X_{n+1}|\theta \sim \text{Bernoulli}(\theta)$. $\underline{X} \perp X_{n+1}|\theta$.

$\theta|X \sim \text{Beta}(\alpha+\sum x_i, \beta+n-\sum x_i)$. $\dfrac{1}{B(\alpha+\beta+n)}\theta^{\alpha+\sum^n x_i}(1-\theta)^{\beta+n-\sum^n x_i}\cdot I_{(0,1)}\theta \cdot I_{\{0,1\}}x_i.$

$$p(X_{n+1}) = \int_{\Omega_\theta} p(X_{n+1}|\theta)\cdot p(\theta)\,d\theta = \int_0^1 \theta^{X_{n+1}}(1-\theta)^{1-X_{n+1}}\cdot\frac{1}{B(\alpha,\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta$$

$$\text{Beta}(\alpha,\beta) = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$= \int_0^1 \theta^{X_{n+1}+\alpha-1}(1-\theta)^{1-X_{n+1}+\beta-1}\frac{1}{B(\alpha,\beta)}\,d\theta = \frac{B(X_{n+1}+\alpha, \beta+1-X_{n+1})}{B(\alpha,\beta)}$$

$$= \frac{\Gamma(\alpha+X_{n+1})\cdot\Gamma(\beta+1-X_{n+1})}{\Gamma(\alpha)\cdot\Gamma(\beta)}\left(\frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha+\beta+1)}}\right) \sim \text{predictive prior dist.}$$

$(\alpha+\beta)$

$$p(X_{n+1}|X) = \int_{\Omega_\theta} p(X_{n+1}|\theta)\,p(\theta|X)\,d\theta = \int_{\Omega_\theta}\theta^{X_{n+1}}(1-\theta)^{1-X_{n+1}}\underline{p(\theta|X)}\,d\theta = \int\frac{1}{B(\alpha+\beta+n)}\theta^{\sum^{n+1}x_i+\alpha-1}(1-\theta)^{(n+1)-\sum^{n+1}x_i+\beta-1}d\theta$$

$\text{Beta}(\alpha+\sum^n x_i, \beta+n-\sum^n x_i)$

$\propto \theta^{\sum x_i+\alpha-1}(1-\theta)^{n-\sum x_i+\beta-1}$

$$R = \frac{B(\alpha+\beta+(n+1))}{B(\alpha+\beta+n)} = \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+\beta+n+1)}\cdot\frac{\Gamma(\alpha+\beta)\Gamma(n+1)\,n}{\Gamma(\alpha+\beta)\cdot\Gamma(1)}$$

$$= \frac{(\alpha+\beta)!}{(\alpha+\beta+n)!}\cdot n = \frac{1}{{}_{\alpha+\beta+n}C_n}\cdot\frac{1}{(n-1)!}.$$

$$= R\int\frac{1}{B(\alpha+\beta+n+1)}\theta^{\sum^{n+1}x_i+\alpha-1}(1-\theta)^{(n+1)-\sum^{n+1}x_i+\beta-1}d\theta.$$

$$= R\int \text{pdf of Beta}(\alpha+\textstyle\sum^{n+1}x_i, \beta+(n+1)-\sum^{n+1}x_i)\,d\theta.$$

⇒ 이렇게 하면 오류 생기는 이유.

※ X_{n+1}은 확률변수고,

$\sum^n x_i$은 given 이고,

$X_{n+1}+\sum^n X_i \neq \sum^{n+1} X_i$ 이다.

$$= \frac{\alpha^{X_{n+1}}\beta^{1-X_{n+1}}}{\alpha+\beta}.$$

$$p(X_{n+1}|X) = \frac{(\alpha+\sum^n x_i)^{X_{n+1}}\cdot(\beta+n-\sum^n x_i)^{1-X_{n+1}}}{\alpha+\beta+n}.$$

we saw $p(\tilde{x}|x)$ when $\tilde{x}|\theta \sim \text{Bernoulli}(\theta)$.

now, consider $\tilde{x}|\theta \sim \text{Bin}(m, \theta)$.

$$p(\tilde{x}|x) = \int_0^1 p(\tilde{x}|\theta)\cdot p(\theta|x)\, d\theta$$

$$= \int_0^1 \binom{m}{\tilde{x}} \theta^{\alpha+\sum x_i+\tilde{x}-1} (1-\theta)^{\beta+n-\sum x_i-\tilde{x}-1} \frac{1}{B(\alpha+\beta+n)} d\theta.$$

$$= \binom{m}{\tilde{x}} \frac{B(\alpha+\beta+n+m+\tilde{x})}{B(\alpha+\beta+n)} \cdot \underbrace{\int_0^1 \frac{1}{B(\alpha+\beta+n+m+\tilde{x})} \theta^{\alpha+\sum x_i+\tilde{x}-1}(1-\theta)^{\beta+n-\sum x_i-\tilde{x}-1} d\theta.}_{\text{sum of pdf Beta}(\alpha+x+\tilde{x},\ \beta+n-x-\tilde{x}) = 1}$$

$\sim$ Beta-Binomial Distribution.

$X \perp \tilde{X} | \theta$.

$X \perp \tilde{X} | \theta. \rightarrow E(\tilde{X}|X) = E[E(\tilde{X}|\theta)|X]$

$\rightarrow Var(\tilde{X}|X) = Var(E(\tilde{X}|\theta)|X) + E(Var(\tilde{X}|\theta)|X)$.

$$E(\tilde{X}|X) = m \cdot \frac{\alpha+\sum x_i}{\alpha+\beta+n}$$

$X \not\perp \tilde{X}$. $X \perp \tilde{X} | \theta$.

코드 구현시 방법.

Method of Composition. $p(\tilde{X}|X) = \int_0^1 p(\tilde{X}|\theta)\cdot p(\theta|X).d\theta$.

[Basic Convergence Assessment with BUGS]

Trace Plots.

ACF.(autocorrelation function plot). : dependence of samples produced by MCMC.

CODA (R).

Detailed code in WinBUGS_betabinom.R. at github.

Laplace's principle of Inverse probability

$$P(E_2|E_1) = \int P(E_2, \theta | E_1)\, d\theta = \int P(E_2 | \theta, E_1) \cdot P(\theta | E_1)\, d\theta$$

$$= \int P(E_2|\theta) \cdot P(\theta|E_1)\, d\theta. \quad (\because E_1 \perp E_2 | \theta).$$ (같이 $\tilde{x}|x$ predictive dist. 할 때 가장 중요한 개념).

(Laplace Rule of Succession). (주요 개념).

JAGS Usage.

Count data: Poisson-Gamma Model.

$X|\theta \sim \mathrm{POI}(\theta)$. $\quad p(x|\theta) = e^{-\theta}\theta^{x}/x!\; I_{\{0,1,2,\dots\}}(x)$. $(\theta > 0)$.

conjugate family for Poisson data: Gamma.

$\theta \sim \mathrm{Gamma}(\alpha, \frac{1}{\beta})$. $\quad p(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)\cdot\beta^{\alpha}}\, \theta^{\alpha-1} e^{-\frac{\beta\theta}{\beta}} I_{(0,\infty)}(\theta)$. $(\alpha > 0, \beta > 0)$.

$\theta | X \sim \mathrm{Gamma}(\alpha + X, (\frac{1}{\beta}+1)^{-1})$ $\frac{\beta}{\beta+1}$. $\quad p(\theta|x) \propto p(x|\theta)\cdot p(\theta) \propto \theta^{(x+\alpha)-1} \cdot e^{-(\frac{1}{\beta}+1)\theta} \cdot I_{(0,\infty)}(\theta)$. $\frac{\beta+1}{\beta}$

theta terms.

$\sim \mathrm{Gamma}(\alpha + X, (\frac{1}{\beta}+1)^{-1})$.

conjugate family (class) definition.

$p(\theta) \in \mathcal{P} \rightarrow p(\theta|x) \in \mathcal{P}$.

$\theta \sim \mathrm{Gamma}(\alpha, \beta)$. $\quad E(\theta|\alpha,\beta) = \frac{\alpha}{\beta}$. $\quad Var(\theta|\alpha,\beta) = \frac{\alpha}{\beta^2}$.

$$E(\theta^k|\alpha,\beta) = \frac{1}{\beta^k}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}. \quad (\alpha+k > 0).$$

코드 구현하는 방법

grid approximation.

Monte Carlo using exact built-in function pdf in R.

Rejection Sampling.

Poisson-Gamma Model.

Lecture Note 6.

$X|\theta \overset{iid}{\sim} POI(\theta)$.

$\theta \sim \text{Gamma}(\alpha, \beta)$. α: shape parameter. : # sum of counts from β prior observations.

β: rate parameter. : # of prior observations.

$\theta|X \sim \text{Gamma}(\alpha + \sum x_i, \beta + n)$

$$E(\theta|x) = \frac{\alpha + \sum x_i}{\beta + n} = \underbrace{\frac{\beta}{\beta+n}}_{\substack{\text{prop. of} \\ \text{prior dist.} \\ p(\theta) \sim \text{Gamma}(\alpha,\beta)}} \cdot \overset{E(\theta)}{\frac{\alpha}{\beta}} + \underbrace{\frac{n}{\beta+n}}_{\substack{\text{prop. of} \\ \text{likelihood.} \\ p(X|\theta) \sim Poi(\theta)}} \cdot \overset{\boxed{E(X|\theta)}?}{\frac{\sum x_i}{n}}.$$ (weighted avg. of prior & likelihood).

Predictive Distribution : NegBin.

$\tilde{X}|\theta \overset{iid}{\sim} \text{Poisson}(\theta)$, $\theta \sim \text{Gamma}(\alpha, \beta)$

- Prior Predictive.

$$p(\tilde{x}) = \int P(\tilde{X} = \tilde{x}|\theta) \cdot P(\theta) \cdot d\theta. = \int P(\tilde{x}|\theta) \cdot P(\theta) \cdot d\theta.$$

$$= \int \frac{e^{-\theta} \cdot \theta^{\tilde{x}}}{\tilde{x}!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha-1} e^{-\beta\theta} I_{(0,\infty)}(\theta) \cdot d\theta = \frac{\beta^\alpha}{\Gamma(\alpha)\,\tilde{x}!} \int \theta^{\alpha+\tilde{x}-1} \cdot e^{-(\beta+1)\cdot\theta} \cdot d\theta.$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha) \cdot \tilde{x}!} \cdot \frac{\Gamma(\alpha+\tilde{x})}{(\beta+1)^{\alpha+\tilde{x}}}. = \binom{\alpha+\tilde{x}-1}{\tilde{x}} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^{\tilde{x}}.$$

$$\sim \text{NegBin}\left(\alpha, \frac{1}{\beta+1}\right).$$

- Posterior Predictive.

$$p(\tilde{x}|x) = \int_0^\infty p(\tilde{x}|\theta) \cdot p(\theta|x).d\theta. = \int POI(\tilde{x}|\theta)\, GAM(\theta|\alpha+\sum x_i, \beta+n)\, d\theta.$$

$$= \frac{\Gamma(\alpha+\sum x_i + \tilde{x})}{\Gamma(\tilde{x}+1)\,\Gamma(\alpha+\sum x_i)} \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum x_i} \left(\frac{1}{\beta+n+1}\right)^{\tilde{x}}.$$

$$\sim \text{NegBin}\left(\alpha+\sum x_i, \frac{1}{\beta+n+1}\right).$$

NegBin.

$$p(X = x|r, p) = \binom{x+r-1}{x} \cdot (1-p)^r \cdot p^x.$$

over-dispersed (more variability) generalization of the Poisson dist.

Exponential Family and Conjugate Priors.

EF $\equiv \phi = \varphi(\theta)$. s.t pdf is in form of

$$p(x|\phi) = h(x) \cdot c(\phi) \cdot \exp(\phi \cdot t(x)). \quad \theta \; c(\phi).$$

any EF has conjugate prior, with pdf

$$p(\phi | n_0, t_0) \propto c(\phi)^{n_0} \exp(n_0 \cdot t_0 \phi). \quad (n_0 > 0, \; t_0 \in \mathbb{R}).$$

resulting posterior is

$$p(\phi | x_1 \cdots x_n) \propto p(\phi) \cdot p(x_1 \cdots x_n | \phi).$$

$$\propto c(\phi)^{n_0+n} \exp\left[\left\{n_0 t_0 + \sum_{i=1}^{n} t(x_i)\right\} \phi\right] \propto p(\phi | n', t').$$

wh. $n' = n_0 + n$.

$$t' = n_0 t_0 + \Sigma\, t(x_i) / n_0 + n. = \frac{n_0}{n_0+n} t_0 + \frac{n}{n_0+n} \cdot \frac{1}{n} \Sigma\, t(x_i).$$

(weighted avg. of prior & likelihood)

$t_0 = E[t(x)]$: prior guess (about distribution). how informative the prior is.

$n_0 =$: number of prior samples.

ex) $X|\theta \sim \text{Bernoulli}(\theta)$.

$$p(x|\theta) = \theta^x (1-\theta)^{1-x} = \frac{e^{\phi x}}{1+e^{\phi}}$$

$$\phi = \ln \frac{\theta}{1-\theta}. \quad e^{\phi} = \left(\frac{1}{1-\theta} - 1\right) \quad \theta = \frac{e^{\phi}}{1+e^{\phi}}$$

(로짓 변환)

$$\theta^x = \frac{e^{\phi x}}{(1+e^{\phi})^x}, \quad (1-\theta)^{1-x} = \frac{1}{(1+e^{\phi})^{1-x}}.$$

$$\theta^x (1-\theta)^{1-x} = \frac{e^{\phi x}}{1+e^{\phi}}.$$

$$c(\phi) = \frac{1}{1+e^{\phi}} = g(\theta).$$ (exp 안에 꼴이 아니고, φ에 대한)

$$p(\theta | n_0, t_0) \propto c(\phi)^{n_0} \exp(n_0 \cdot t_0 \phi).$$

$$\propto \left(\frac{1}{1+e^{\phi}}\right)^{n_0} \cdot e^{n_0 \cdot t_0 \phi}$$

$$p(\theta | n_0, t_0) = p(\phi | n_0, t_0) \cdot \left|\frac{d\phi}{d\theta}\right|. \propto (1-\theta)^{n_0} \left(\frac{\theta}{1-\theta}\right)^{n_0 t_0} \left|\frac{1}{\theta(1-\theta)}\right| = \theta^{n_0 t_0 - 1} (1-\theta)^{n_0(1-t_0)-1}$$

$\phi \to \theta$ 변환.

$\sim \text{Beta}(n_0 t_0, \; n_0(1-t_0)).$

$$\left(\begin{aligned} &\phi = \ln \frac{\theta}{1-\theta} = \ln f(\theta). \quad f(\theta) = 1 - \frac{1}{1-\theta}. \quad f'(\theta) = +\frac{1}{(1-\theta)^2} \\ &\frac{d\phi}{d\theta} = \frac{d \ln f(\theta)}{d\theta} = \frac{f'(\theta)}{f(\theta)} = \frac{\cancel{1-\theta}}{\theta} \cdot \frac{1}{(1-\theta)^{\cancel{2}}} = \frac{1}{\theta(1-\theta)}. \\ &e^{\phi} = \frac{\theta}{1-\theta}. \quad c(\phi) = 1-\theta \end{aligned}\right)$$