Seoul National University
Dept. of Civil and Environmental Engineering

Instructor: Junho Song
junhosong@snu.ac.kr

457.212 Statistics for Civil & Environmental Engineers
In-Class Material: Class 26 (Final)
Introduction to Bayesian Approach (A&T: 9.1-9.3)

Given: Sample data set $\{x_1, x_2, \ldots, x_n\}$ assumed to follow a distribution model $f_X(x;\theta)$

Question: Considering the distribution parameter θ as a *random variable*, what is $f_\Theta(\theta)$?

→ Bayesian parameter estimation

1. **Bayes Rule and Bayesian Parameter Estimation**

(a) Bayes rule (Thomas Bayes, 1763)

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A)$$

Update our degree of b_______ on the event A based on new evidence B

(b) Bayesian parameter estimation?

In non-Bayesian approaches, the unknown parameters of a probability distribution, e.g. $\mu, \sigma, \lambda, \zeta \ldots$ are considered d____________ quantities.

The uncertainty in their estimation has been acknowledged, but in terms of a ________ interval.

The Bayesian approach: any uncertainty quantity is treated as a __________ __________ → Rules of p________ are used to assess and analyze *all* types of uncertainties, i.e. inherent randomness, model and measurement errors, and statistical uncertainties.

Following the Bayesian approach, how to find the distribution of the unknown/uncertain parameter θ in $f_X(x;\theta)$, i.e. $f_\Theta(\theta)$ based on samples $\{x_1, x_2, \ldots, x_n\}$?

2. **Bayesian Parameter Estimation**

(a) Bayesian updating formula

Let us define the events appearing in Bayes rule as follows:

- Target event: $A = \{\theta < \Theta \leq \theta + d\theta\}$
- Evidence: $B = \bigcap_{i=1}^{n} E_i$

Now, the Bayes rule becomes

$$P(\theta < \Theta \leq \theta + d\theta | \cap_{i=1}^{n} E_i) = \frac{P(\cap_{i=1}^{n} E_i \,|\theta < \Theta \leq \theta + d\theta)}{P(\cap_{i=1}^{n} E_i)} P(\theta < \Theta \leq \theta + d\theta)$$

Using the PDF definition,

$$f_\Theta(\theta | \cap_{i=1}^{n} E_i) d\theta = \frac{P(\cap_{i=1}^{n} E_i | \Theta = \theta)}{P(\cap_{i=1}^{n} E_i)} f_\Theta(\theta) d\theta$$

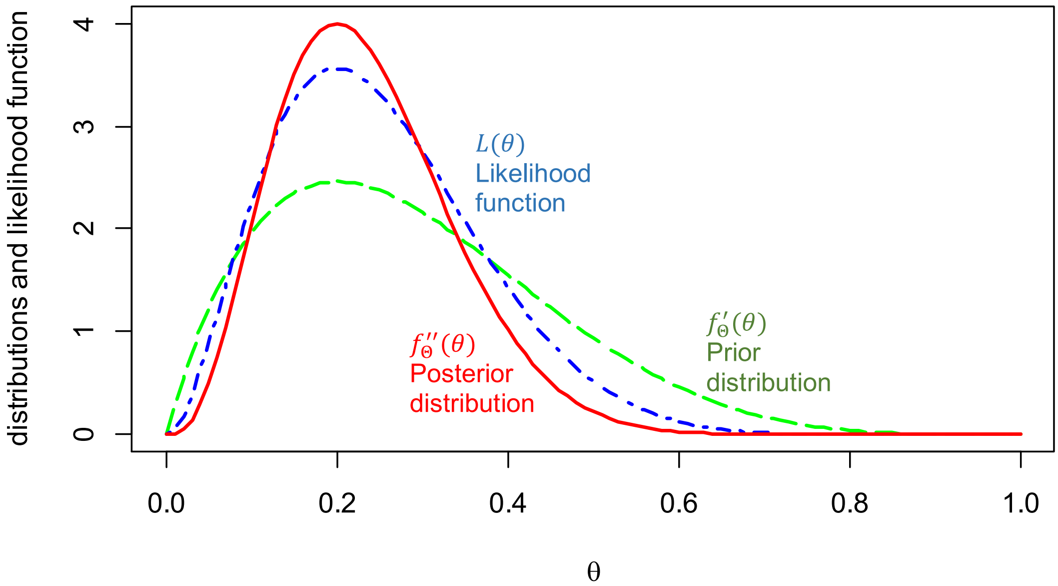
Dividing by $d\theta$ and introducing new notations, we obtain the Bayesian updating formula

$$f''_\Theta(\theta) = c \cdot L(\theta) \cdot f'_\Theta(\theta)$$

where

- $L(\theta) \propto P(\cap_{i=1}^{n} E_i | \Theta = \theta)$ is "Likelihood Function," e.g. $\prod_{i=1}^{n} f_X(x_i; \theta)$
- $f'_\Theta(\theta)$ is the "Prior Distribution" ~ Degree of belief before $\cap_{i=1}^{n} E_i$ occurs (or assumed)
- $f''_\Theta(\theta)$ is the "Posterior Distribution" ~ Updated degree of belief after $\cap_{i=1}^{n} E_i$
- c is the normalizing factor (to make $f''_\Theta(\theta)$ a valid PDF)

$$c = \left[\int_{-\infty}^{\infty} L(\theta) f'_\Theta(\theta) d\theta \right]^{-1}$$



- The sharpness of the curve indicates the information content: the sharper, the more information
- The posterior distribution is generally _________ than the other two because it contains the information content of both curves
- The posterior statistics can be computed using $f''_\Theta(\theta)$, e.g.

$$\mu''_\theta = \int_{-\infty}^{\infty} \theta f''_\Theta(\theta) d\theta$$

$$(\sigma''_\theta)^2 = \int_{-\infty}^{\infty} (\theta - \mu''_\theta)^2 f''_\Theta(\theta) d\theta$$

- It is generally challenging to compute c, μ''_θ and σ''_θ, especially when $\theta \rightarrow \boldsymbol{\theta}$

Instructor: Junho Song
junhosong@snu.ac.kr

(b) "Predictive" distribution $\tilde{f}_X(x)$

When the inherent variability in X and the uncertainty in θ are combined, the "predictive" distribution of X is given as

$$\tilde{f}_X(x) = \int_{-\infty}^{\infty} f_X(x;\theta) f''_\Theta(\theta) d\theta$$

(c) Choice of the prior distribution $f'_\Theta(\theta)$

The prior distribution reflects the state of knowledge about the parameter θ before observations are made. There are various available options including:

- **Subjective Prior**: a distribution reflecting expert judgment and opinion. Determine mean and standard deviation, and then select a type of distribution considering the boundedness and/or possibility of using conjugate priors
- **Previous Posterior as Prior**: the posterior distribution from a previous updating. (Note: sequential updating should give the same result as an updating based on the all data sets; the sequence does not affect the end result either)
- **Diffuse Prior**: Flat distribution in the non-zero interval to avoid influence on the posterior
- **Non-informative Prior**: prior distribution reflecting no prior information regardless of the parameter transformation $\eta = \eta(\theta)$ (Box and Tiao, 1992)

3. Conjugate Prior Distributions

(a) Recall the Bayesian updating formula:

$$f''_\Theta(\theta) = c \cdot L(\theta) \cdot f'_\Theta(\theta)$$

If the use of a certain type of distribution for $f'_\Theta(\theta)$ guarantees that the posterior distribution $f''_\Theta(\theta)$ follows the same distribution, we call the type of distribution as "____________ ___________" distribution for the sample distribution $f_X(x;\theta)$

(b) In such cases, we do not need to compute c, μ''_θ and σ''_θ by numerical integrations. Just obtain the posterior statistics by closed-form formulas

Example 1: Conjugate prior $f'_\Theta(\theta)$ of the Binomial distribution $f_X(x;\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$?

Likelihood function: $L(\theta) =$

To be a conjugate prior, the prior distribution should keep the same mathematical form after being multiplied by $L(\theta)$

Let us try the Beta distribution as $f'_\Theta(\theta)$, i.e.

$$f'_\Theta(\theta) = \frac{\Gamma(q'+r')}{\Gamma(q')\Gamma(r')}\theta^{q'-1}(1-\theta)^{r'-1}, 0 < \theta < 1$$

According to the Bayesian updating formula,

$$f''_\Theta(\theta) = c \qquad \frac{\Gamma(q'+r')}{\Gamma(q')\Gamma(r')}\theta^{q'-1}(1-\theta)^{r'-1}$$
$$\propto \theta^{q''-1}(1-\theta)^{r''-1}$$

where

$$q'' =$$
$$r'' =$$

It is seen that the posterior distribution is beta with its parameters q'' and r'' obtained by the closed-form updating formula given above.

The corresponding predictive distribution is derived as so-called Beta-Binomial distribution

$$\tilde{f}_X(x) = \int_0^1 f_X(x;\theta) f''_\theta(\theta) d\theta$$
$$= \frac{n!}{x!\,(n-x)!}\frac{\Gamma(q''+r'')}{\Gamma(q'')\Gamma(r'')}\frac{\Gamma(q''+x)\Gamma(r''+n-x)}{\Gamma(q''+r+n)}$$

(c) Extensive table of conjugate distributions is available, e.g.,
https://en.wikipedia.org/wiki/Conjugate_prior

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters	Interpretation of hyperparameters[note 1]	Posterior predictive[note 2]
Bernoulli	p (probability)	Beta	α, β	$\alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i$	$\alpha - 1$ successes, $\beta - 1$ failures[note 1]	$p(\tilde{x}=1) = \frac{\alpha'}{\alpha'+\beta'}$
Binomial	p (probability)	Beta	α, β	$\alpha + \sum_{i=1}^n x_i, \beta + \sum_{i=1}^n N_i - \sum_{i=1}^n x_i$	$\alpha - 1$ successes, $\beta - 1$ failures[note 1]	$\mathrm{BetaBin}(\tilde{x}\mid\alpha',\beta')$ (beta-binomial)
Negative binomial	p (probability)	Beta	α, β	$\alpha + \sum^n x_i, \beta + rn$	$\alpha - 1$ total successes, $\beta - 1$ failures[note 1]	

Other notable conjugate priors: Gamma for Poisson, Normal for Normal (with known variance), etc.

4. Applications of Bayesian Parameter Estimation

(a) Estimation with "perfect" tests (Der Kiureghian 2009)

Consider a quarry with a mixture of "good" and "bad" materials, and let θ denote the fraction of good materials in the mixture.

Instructor: Junho Song
junhosong@snu.ac.kr

Suppose the number of good samples out of n tested samples, denoted by X follows the binomial distribution.

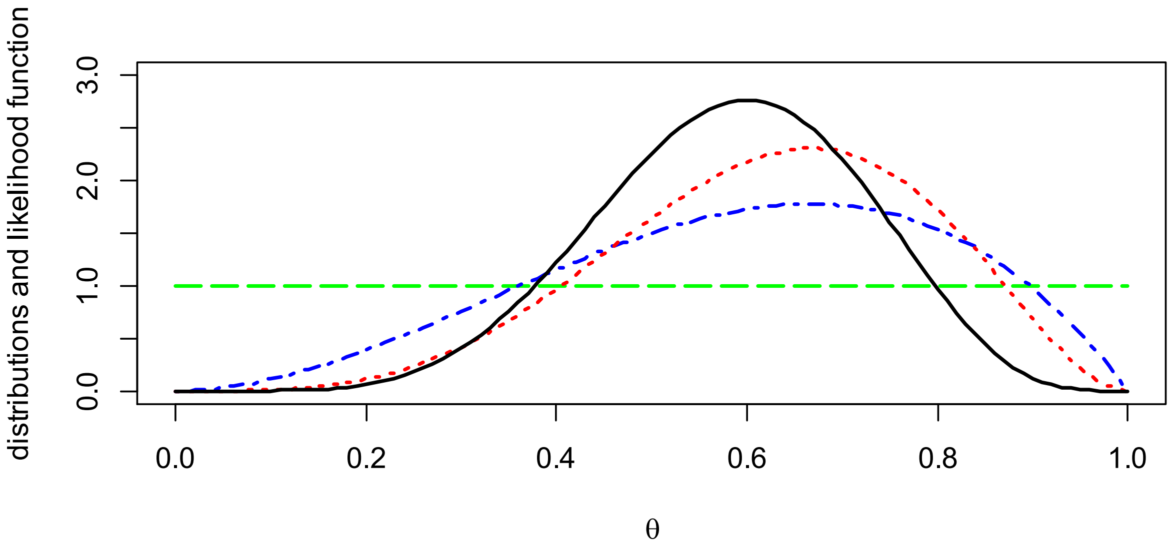
Example 1 showed that the conjugate prior of θ is Beta distribution.The posterior distribution has hyper-parameters $q'' = q' + x$ and $r'' = r' + n - x$.

From the property of Beta distribution, the posterior mean and c.o.v. are

$$\mu''_\theta = \frac{q''}{q'' + r''}$$

$$\delta''_\theta = \sqrt{\frac{r''}{q''(q'' + r'' + 1)}}$$

Suppose the prior distribution is assumed to be Uniform, i.e. Beta with $q' = 1$ and $r' = 1$. The following test results were obtained sequentially: (i) 2 out of 3 samples are good, (ii) 2 out of next 3 samples are good, (iii) 2 out of next 4 samples. obtain the posterior PDFs and the corresponding mean and c.o.v.




```
# (2) Application - Perfect test

rm(list=ls())
th = seq(0,1,0.01)
q=1
r=1
p = dbeta(th,shape1=q,shape2=r) # conjugate prior - beta (start with
uniform)
mu0=q/(q+r)
cov0=sqrt(r/q/(q+r+1))

# 1st test: n=3, x=2
q=q+2
r=r+3-2
f1=dbeta(th,shape1=q,shape2=r)
mu1=q/(q+r)
```

Instructor: Junho Song
junhosong@snu.ac.kr

```
cov1=sqrt(r/q/(q+r+1))

# 2nd test: n=3, x=2
q=q+2
r=r+3-2
f2=dbeta(th,shape1=q,shape2=r)
mu2=q/(q+r)
cov2=sqrt(r/q/(q+r+1))

# 3rd test: n=4, x=2
q=q+2
r=r+4-2
f3=dbeta(th,shape1=q,shape2=r)
mu3=q/(q+r)
cov3=sqrt(r/q/(q+r+1))

plot(th,p,xlab=expression(theta),ylab="distributions and likelihood
function",type="l",lty=5,ylim=c(0,3),col="green",lwd=2)
lines(th,f1,type="l",lty=4,col="blue",lwd=2)
lines(th,f2,type="l",lty=3,col="red",lwd=2)
lines(th,f3,type="l",lty=1,col="black",lwd=2)

post_means = c(mu0,mu1,mu2,mu3)
post_covs = c(cov0,cov1,cov2,cov3)
```

(b) Estimation with "imperfect" tests (Der Kiureghian 2009)

Suppose each test is not perfect, i.e. could give a wrong outcome from the test.

Test outcome	True state is Good (G)	True state is Bad (B)
Test indicates Good (IG)	0.7	0.1
Test indicates Bad (IB)	0.1	0.8
Test is inconclusive (IN)	0.2	0.1

Likelihood function $\propto P(x\text{ "IG"}, y\text{ "IB"}, \text{and } (n-x-y)\text{"IN"}|\Theta=\theta)$

$$P(IG) = P(IG|G)P(G) + P(IG|B)P(B) = 0.7\theta + 0.1(1-\theta) = 0.1 + 0.6\theta$$
$$P(IB) = P(IB|G)P(G) + P(IB|B)P(B) = 0.1\theta + 0.8(1-\theta) = 0.8 - 0.7\theta$$
$$P(IN) = P(IN|G)P(G) + P(IN|B)P(B) = 0.2\theta + 0.1(1-\theta) = 0.1 + 0.1\theta$$

Therefore, the likelihood function is

$$L(\theta) = (0.1+0.6\theta)^x(0.8-0.7\theta)^y(0.1+0.1\theta)^{n-x-y}$$

No more conjugate priors! We need to directly rely on the Bayesian updating formula

$f''_\Theta(\theta) = cL(\theta)f'_\Theta(\theta)$ with

$$c = \left[\int_{-\infty}^{\infty} L(\theta)f'_\Theta(\theta)d\theta\right]^{-1}$$
$$= \left[\int_{-\infty}^{\infty} (0.1+0.6\theta)^x(0.8-0.7\theta)^y(0.1+0.1\theta)^{n-x-y} f'_\Theta(\theta)d\theta\right]^{-1}$$

Instructor: Junho Song
junhosong@snu.ac.kr

For a single parameter θ, you could do this by numerical integration, but could be challenging if there are more than one parameter to estimate.

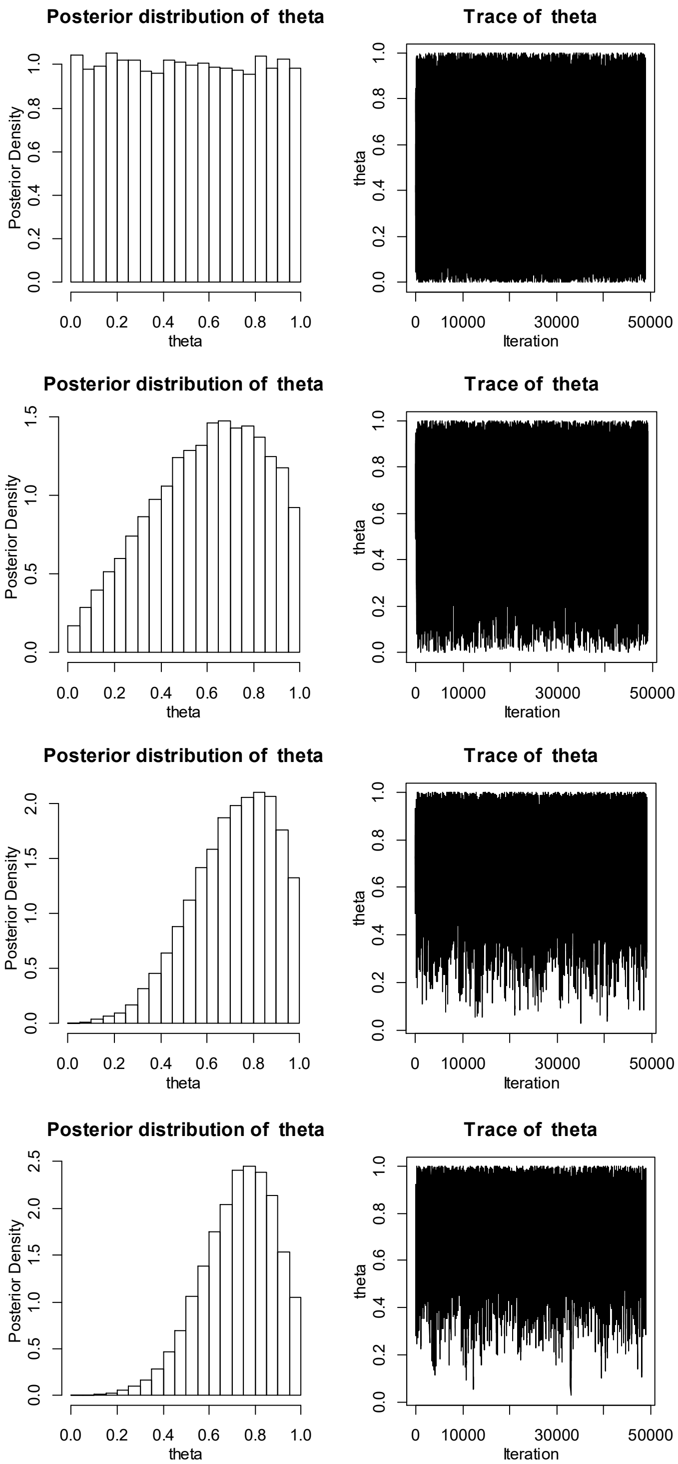
Researchers had developed special algorithms for the purpose, but recently, we had so-called … MCMC revolution (Diaconis, 2009)!

Although this is a single-parameter estimation, let us perform Bayesian parameter estimation by a Markov Chain Monte Carlo method (Metropolis-Hastings algorithm).

```
# (3) Bayesian parameter estimation by MCMC
# Ref: https://bayesianbiologist.com/2012/02/06/general-bayesian-estimation-
using-mhadaptive/
rm(list=ls())
install.packages("MHadaptive")
library(MHadaptive)

li_est=function(pars,data) #likelihood function
{
  theta = pars[1]
  n = data[1]
  x = data[2]
  y = data[3]
  likelihood = (0.1+0.6*theta)^x*(0.8-0.7*theta)^y*(0.1+0.1*theta)^(n-x-y)
  log_likelihood = log(likelihood)
  prior = prior_est(pars)
  return(log_likelihood + prior)
}

prior_est=function(pars) # prior distribution
{
  theta = pars[1]
  prior=dunif(theta,0,1,log=TRUE)
  return(prior)
}

m0=Metro_Hastings(li_func=li_est,pars=c(0.5),par_names=c('theta'),data=c(0,0,0))
m1=Metro_Hastings(li_func=li_est,pars=c(0.5),par_names=c('theta'),data=c(3,2,1))
m2=Metro_Hastings(li_func=li_est,pars=c(0.5),par_names=c('theta'),data=c(8,5,2))
m3=Metro_Hastings(li_func=li_est,pars=c(0.5),par_names=c('theta'),data=c(12,7,3))
# c(3,2,1) means n=3, x=2, y=1 (2 "Good", 1 "Bad", 0 “Inconclusive”)

plotMH(m0)
plotMH(m1)
plotMH(m2)
plotMH(m3)

PostMean=c(mean(m0$trace),mean(m1$trace),mean(m2$trace),mean(m3$trace))
PostSd=c(sd(m0$trace),sd(m1$trace),sd(m2$trace),sd(m3$trace))
PostCOV = PostSd/PostMean
```

Histograms representing the prior and the posterior distributions after three tests:

(1) $n = 3, x = 2, y = 1$ (2 “IG”, 1 “IB”, 0 “IN”)
(2) $n = 5, x = 3, y = 1$ (3 “IG”, 1 “IB”, 1 “IN”)
(3) $n = 4, x = 2, y = 1$ (2 “IG”, 1 “IB”, 1 “IN”)



“The best thing about being a statistician is that you get to play in everyone’s backyard.” – John Tukey (1915-2000)

“What would be the best thing about knowing probability and statistics as a civil and environmental engineer? I hope my humble course helped you get ready to answer the question in near future. Thank you♥” – J.S.