Theorem 10.6 (2nd Ed., TBp.590-591) cf. Thm10.5 (LNp.21-22)

- Consider the data $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, where — dist. of data in Frequentist; cond'nal dist. of data in Bayesian
  - sample size
  - $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where μ is parameter and σ^2 known
  - the prior distribution of μ is $N(\mu_0, \sigma_0^2)$. — assumed known
- Let $\xi = 1/\sigma^2$ ($X|\theta$) and $\xi_0 = 1/\sigma_0^2$ (θ).
- The posterior distribution of μ is normal with mean

$$\mu_1 = \frac{\xi_0\mu_0 + n\xi\bar{x}}{n\xi + \xi_0} = \frac{\xi_0}{n\xi + \xi_0}\mu_0 + \frac{n\xi}{n\xi + \xi_0}\bar{x}$$

Bayes est'or of μ → μ_1 — biased est'or under $X|\mu$

Best estimate of μ before observing data (μ_0); Frequentist estimator of μ ($\bar{x}$); if $\xi \ll \xi_0$, if $\xi_0 \ll \xi$; sum $=1$; if n large, if n small

every observation carries a certain amount of precision

and precision $\xi_1 = n\xi + \xi_0$ (i.e., variance $1/\xi_1$). — $< 1/\xi_0 = \sigma_0^2$

Proof. The posterior distribution of μ is

$$h(\mu|\mathbf{x}) \propto f(x_1, \ldots, x_n|\mu)\, g(\mu) \propto \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \mu)^2\right] \times \exp\left[-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right]$$

$$\propto \exp\left[-\frac{1}{2a^{-1}}\left(\mu - \frac{b}{a}\right)^2\right],$$

where $a = n\xi + \xi_0$, and $b = \xi_0\mu_0 + n\xi\bar{x}$.

cf. Fisher Information; It's a pdf of μ; $\sum_{i=1}^{n}(x_i - \underbrace{\mu}_{-\bar{x}+\bar{x}})^2$; similar form for μ

Notes (2nd Ed., TBp.591).

For large sample size n, $\mu_1 \approx \bar{x}$, $\xi_1 \approx n/\sigma^2$, → posterior $\approx$ likelihood $N(\bar{x}, \sigma^2/n)$ ∝ — cf. item5, LNp.23

$\because n\xi \gg \xi_0$

posterior ∝ likelihood × prior

prior: minor influence; likelihood: dominant

the information in the sample (data) largely determines the posterior distribution.

Example 10.8 (2nd Ed., TBp. 591-592) μ (r.v.) $\sim N(2, 4) \Rightarrow$ 4: a realization of μ — unknown

- Suppose prior of μ is $N(2, 4)$ and $X_1, X_2, \ldots, X_n$ are i.i.d. from $N(4, 1)$. — cf. true parameter; but, not observed
  - $n = 2$, $x_1 = 3.59$, $x_2 = 5.52$,

    update → posterior: $\mu_1 = 0.89\,\bar{x}_2 + 0.11\,\mu_0 = 4.27$, $\xi_1 = 2.25$ — $\mu|X_1, X_2$
  - prior, update: $n = 4$, $x_3 = 3.93$, $x_4 = 4.71$,

    posterior: $\mu_1 = 0.94\,\bar{x}_4 + 0.06\,\mu_0 = 4.29$, $\xi_1 = 4.25$ — $\mu|X_1, X_2, X_3, X_4$
  - prior, update: $n = 8$, $x_5 = 4.40$, $x_6 = 5.06$, $x_7 = 3.68$, $x_8 = 3.14$,

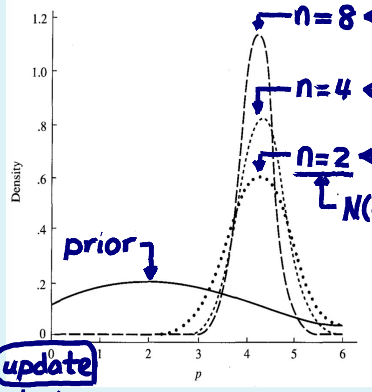
    posterior: $\mu_1 = 0.97\,\bar{x}_8 + 0.03\,\mu_0 = 4.19$, $\xi_1 = 8.25$ — update, posterior; $\mu|X_1, \cdots X_8$

  Frequentist estimator ($\bar{x}_8$)

  $f(\mu|x_1, \cdots, x_8) \propto f(x_5, \cdots, x_8|\mu) \cdot g(\mu|x_1, \cdots, x_4)$

  $\because$ Conditioned on μ, $X_1, \cdots, X_8$ are independent

FIGURE 15.4 The prior density of μ (solid line) and the posterior densities after 2 (dotted line), 4 (short-dashed line), and 8 (long-dashed line) observations.



- **Figure 15.4**:
  - Prior density of μ (solid), and posterior density of $\mu|\mathbf{x}$ after 2 (dot), 4 (short-dash), and 8 (long-dash) observations.
  - The prior density is quickly dominated by the data.
  - The posterior density becomes more concentrated, at the true mean 4, as the sample size increases.

❖ **Reading**: textbook (2nd ed.), 15.3.1

- Bayesian Inference for the Binomial distribution — $X|\theta$

Theorem 10.7 (2nd Ed., TBp. 593-594)

Consider the data

- $X \sim \text{Binomial}(n, p)$, i.e., $f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, \ldots, n$

sample size; random; fixed; $0 < p < 1$, parameter; dist. of data in Frequentist; cond'nal dist. of data in Bayesian

- prior distribution of p is Beta(a, b), i.e., $g(p) \propto p^{a-1}(1-p)^{b-1}$, $0 \le p \le 1$

assumed known for simplicity; similiar form for p; random

Then, the posterior distribution of p given $X = x$ is Beta$(\underbrace{a+x}_{a'}, \underbrace{n+b-x}_{b'})$, i.e.,

$$h(p\,|\,x) \propto p^{\overbrace{a+x}^{a'}-1}(1-p)^{\overbrace{n+b-x}^{b'}-1}$$

update

Proof. The posterior distribution of p is

This is a pdf of p

$$h(p\,|\,x) \propto f(x\,|\,p)\,g(p) \propto \left[p^x(1-p)^{n-x}\right]\cdot\left[p^{a-1}(1-p)^{b-1}\right]$$

$$= p^{\overbrace{a+x}^{a'}-1}(1-p)^{\overbrace{n+b-x}^{b'}-1}$$

biased estimator under $X|p$ (admissible)

if n large; if n small

Notes (2nd Ed., TBp. 594).

- The posterior mean is

$Z \sim \text{Beta}(\alpha, \beta)$, $E(Z) = \frac{\alpha}{\alpha+\beta}$ (LN, CH1~6, p.75)

Bayes est'or of p; best estimate of p before observing data; Frequentist estimator of p

$$\mu_{\text{post}} = \frac{a+x}{a+b+n} = \frac{a+b}{a+b+n}\cdot\frac{a}{a+b} + \frac{n}{a+b+n}\cdot\bar{x}, \quad \bar{x} = \frac{x}{n}$$

sum = 1

which is a weighted average of the prior mean $\frac{a}{a+b}$ and sample mean $\bar{x}$.

- If p_0 is the true value of parameter p, then

realization of r.v. p; unknown; (LN, CH1~6, p. 3)

$$\bar{x} \xrightarrow{P} p_0, \quad \text{as } n \to \infty. \leftarrow \text{by Law of Large Number}$$

Similarly,

$$\mu_{\text{post}} \xrightarrow{P} p_0, \quad \text{as } n \to \infty. \Rightarrow \text{weights} \begin{cases} \frac{a+b}{a+b+n} \to 0 \\ \frac{n}{a+b+n} \to 1 \end{cases}$$

What if $a+b \gg n$? What does it mean?

subjective probabilities (Bayesian) $\downarrow n\to\infty$ objective probability (Frequentist) (check *, LN p.19)

- The posterior variance is

$$\frac{(a+x)(n+b-x)}{(a+b+n)^2(a+b+n+1)},$$

posterior $\propto$ likelihood

$Z \sim \text{Beta}(\alpha, \beta)$, $Var(Z) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$ (LN, CH1~6, p.73)

which tends to zero as $n \to \infty$, i.e., the posterior becomes more and more concentrated about p_0.

- When $a = b = 1$, the prior of p is Uniform$(0, 1)$, which expresses an indifference about the possible values of p. noninformative prior

Example 10.9 (2nd Ed., TBp. 596)

Q: Which data, X or Y, is more informative? Q: Is it always better to draw inference based on Y?

- (*Thomas, 1948*) A cofferdam protecting a construction site was designed to withstand flows of up to 1870 ft^3/sec (cfs).

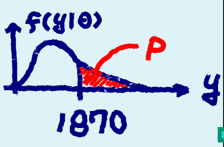
Q: What if we use this data to peform the analysis?

- An engineer wishes to compute the probability p that the dam will be overtopped during the coming year.

(continuous) $Y_1, \cdots, Y_{25} \sim f(y|\theta)$

- discrete: Data X. Over the past 25 years, the annual maximum flood levels had ranged from 629 to 4720 cfs, and 1870 ft^3/sec had been exceeded in $x = 5$ out of the past 25 years.

$X_i = I_{(1870,\infty)}(Y_i)$



- Statistical modeling: assume a constant p over years → What if not
  - model the 25 years as 25 independent Bernoulli trials $X_i \sim B(p)$, $i = 1, \ldots, 25$. Then, $X = \sum_{i=1}^{25} X_i \sim \text{Binomial}(25, p)$

  Alternative modeling: (X_i, P_i), $i=1, \ldots, 25$, are i.i.d. with $X_i | P_i \sim B(P_i)$, $P_i \sim U(0,1)$

  Bayesian approach: posterior ∝ likelihood

  - update — use a Uniform(0, 1) prior for p ← Beta(1,1) — noninformative
- the posterior distribution is

$$\text{Beta}(6,21) \leftarrow h(p\,|\,x) \propto p^x(1-p)^{n-x} = p^{6-1}(1-p)^{21-1}$$

small weight, large weight

- posterior mean $= \frac{2}{2+25} \cdot \frac{1}{2} + \frac{25}{2+25} \cdot \frac{5}{25} = \frac{6}{27}$

$\frac{5}{25}$: Frequentist estimate; $\frac{6}{27}$: Bayesian estimate (cf.) — different even though the prior pdf is a constant

Definition 10.5 (conjugate priors, 2nd Ed., TBp. 596-597)

Why need this?

- G: family of prior distributions $g(\theta)$ for Θ ← cf.
- H: family of conditional distribution $f(\mathbf{x}|\theta)$ of data $(\mathbf{X})$ given $\Theta = \theta$ → The joint distribution of data in Frequentist approach

G is called a family of **conjugate priors** to H if the posterior distribution also belongs to G. → i.e., prior and posterior belong to "same type" of distribution.

Example 1. Normal prior for the mean μ of a Normal data (μ: parameter)

Thm10.6 (LNp.24) H: normal dist. G: normal dist. → posterior is normal

Example 2. Beta prior for the probability p of success in a Binomial data (p: parameter)

Thm10.7 (LNp.26) H: binomial dist. G: Beta dist. → posterior is Beta

❖ **Reading**: textbook (2nd ed.), 15.3.2