

Suggested Title: Chapter 7 _____ [Edit](#)

14 November 2019, 8:15

Chapter 7 Solutions

① $w \sim N(M, A)$

$x|w \sim N(\mu, D)$, $D > 0$ known.

a) What is the marginal distribution of x ?

• First, we write the joint distribution $f(x, w)$ and use conditional probability to write:

$$f(x, w) = f(x|w) \cdot f(w)$$



use conditional probability

$$f(x, \omega) = f(x|\omega) \cdot f(\omega)$$

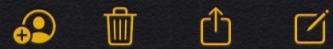
Further, in order to compute $f(x)$, the marginal distribution of x , we need to compute:

$$f(x) = \int_{-\infty}^{\infty} f(x, \omega) d\omega = \int_{-\infty}^{\infty} f(x|\omega) \cdot f(\omega) d\omega$$

Thus, using the Normal pdf.

$$f(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left\{-\frac{1}{2} \cdot \left[\frac{(x-\mu)^2}{\sigma^2}\right]\right\} \cdot \frac{1}{\sqrt{2\pi A}} \cdot \exp\left\{-\frac{1}{2} \cdot \left[\frac{(\omega-\mu)^2}{A}\right]\right\} \cdot d\omega$$

... write the integrand



Now, our strategy will be to write the integrand as a term that does not include μ (and thus can go out of the integral) and a Normal kernel of the form $\exp\left\{-\frac{1}{2} \cdot \frac{(\mu - E(\mu))^2}{V(\mu)}\right\}$, since we know that

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \cdot \frac{(\mu - E(\mu))^2}{V(\mu)}\right\} d\mu = \frac{1}{\sqrt{2\pi \cdot V(\mu)}}. \text{ Ok, let's start...}$$

$$f(x) = \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \cdot \left[\frac{(x-\mu)^2}{D} + \frac{(\mu-M)^2}{A}\right]\right\} d\mu$$

$$= \frac{1}{\sqrt{A}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \cdot \left[\frac{x^2 - 2\mu x + \mu^2}{D} + \frac{\mu^2 - 2\mu M + M^2}{A}\right]\right\} d\mu$$

$$\begin{aligned}
 f(x) &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{(x-\mu)^2}{D} + \frac{(\mu-M)^2}{A} \right] \right\} \cdot d\mu \\
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{x^2 - 2\mu x + \mu^2}{D} + \frac{\mu^2 - 2\mu M + M^2}{A} \right] \right\} \cdot d\mu \\
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{Ax^2 - 2\mu Ax + A\mu^2 + D\mu^2 - 2\mu DM + Dm^2}{AD} \right] \right\} \cdot d\mu \\
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{\mu^2(A+D) - 2\mu(Ax+DM)}{AD} + \frac{Ax^2 + DM^2}{AD} \right] \right\} \cdot d\mu \\
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{\mu^2 - 2 \cdot \frac{\mu(Ax+DM)}{(A+D)}}{AD/(A+D)} + \frac{Ax^2 + DM^2}{AD} \right] \right\} \cdot d\mu
 \end{aligned}$$



$$= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{w^2 - 2 \cdot \frac{w(Ax+Dm)}{(A+D)}}{AD/(A+D)} + \frac{Ax^2 + Dm^2}{AD} \right] \right\} dw$$

Now, to form the Normal kernel we wanted, we need to complete the square of $w^2 - 2 \cdot w \cdot \left(\frac{Ax+Dm}{(A+D)} \right)$. Thus,

$$= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{w^2 - 2 \cdot w \left(\frac{Ax+Dm}{(A+D)} \right) + \left(\frac{Ax+Dm}{(A+D)} \right)^2}{AD/(A+D)} + \frac{Ax^2 + Dm^2}{AD} - \left(\frac{Ax+Dm}{(A+D)} \right)^2 \cdot \frac{(A+D)}{AD} \right] \right\} dw$$

$$= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{\left(w - \left(\frac{Ax+Dm}{(A+D)} \right) \right)^2}{AD/(A+D)} + \frac{Ax^2 + Dm^2}{AD} - \frac{(Ax+Dm)^2}{(A+D) \cdot AD} \right] \right\} dw$$

$$= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \cdot \left[\frac{\left(w - \left(\frac{Ax+Dm}{(A+D)} \right) \right)^2}{AD/(A+D)} + \frac{Ax^2 + Dm^2 + ADx^2 + D^2m^2 - A^2x^2 - 2AxDm - D^2m^2}{AD} \right] \right\} dw$$



$$\begin{aligned}
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left[\frac{w - \frac{(Ax+Dm)}{(A+D)}}{\sqrt{AD/(A+D)}} \right]^2 - \frac{AD}{(A+D)} \right\} d\mu \\
 &\quad \text{Note: } \frac{1}{2} \left[\frac{(w - \frac{(Ax+Dm)}{(A+D)})^2}{AD/(A+D)} + \frac{A^2x^2 + ADm^2 + ADx^2 + D^2m^2 - A^2x^2 - 2AxDm - D^2m^2}{AD} \right] \\
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left[\frac{(w - \frac{(Ax+Dm)}{(A+D)})^2}{AD/(A+D)} + x^2 - 2xm + m^2 \right] \right\} d\mu \\
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left[\frac{(w - \frac{(Ax+Dm)}{(A+D)})^2}{AD/(A+D)} + (x-m)^2 \right] \right\} d\mu \\
 &= \frac{1}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left[\frac{(w - \frac{(Ax+Dm)}{(A+D)})^2}{AD/(A+D)} + (x-m)^2 \right] \right\} d\mu \\
 &= \frac{\exp\left\{-\frac{1}{2} \cdot \frac{(w - \frac{(Ax+Dm)}{(A+D)})^2}{AD/(A+D)}\right\}}{2\pi\sqrt{AD}} \cdot \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \cdot \frac{(x-m)^2}{AD/(A+D)}\right\} d\mu
 \end{aligned}$$

$$\begin{aligned}
 & \text{EITVAU} \quad \text{---} \quad AD / (A+D) \\
 & = \frac{\exp\left\{-\frac{1}{2} \cdot [(X-m)^2]\right\}}{2\pi \sqrt{AD}} \cdot \underbrace{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \cdot \left[\frac{(w - \left(\frac{Ax + Dm}{A+D}\right))^2}{AD / (A+D)}\right]\right\} dw}_{N\left(\frac{Ax + Dm}{A+D}, \frac{AD}{A+D}\right) \text{ kernel.}} \\
 & = \frac{\exp\left\{-\frac{1}{2} \cdot [(X-m)^2]\right\}}{2\pi \sqrt{AD}} \cdot \frac{1}{\sqrt{2\pi} \cdot \sqrt{\frac{AD}{A+D}}} \\
 & = \frac{\exp\left\{-\frac{1}{2} \cdot [(X-m)^2]\right\}}{\text{---} \quad \text{---} \quad \text{---}}
 \end{aligned}$$



$$\cdot = \frac{\exp\left(-\frac{1}{2} \cdot ((x-m)^2)\right)}{(2\pi)^{3/2} \cdot \left(\frac{AD}{\sqrt{A+D}}\right)}$$

$$\Rightarrow f(x) = \frac{1}{(2\pi)^{3/2} \cdot \left(\frac{AD}{\sqrt{A+D}}\right)} \cdot e^{-\frac{1}{2} \cdot (x-m)^2}$$

yay!

b) We can get $f(\mu|x)$ using Bayes Rule:

$$f(\mu|x) = \frac{f(x|\mu)}{f(x)} = \frac{f(x|\mu) \cdot f(\mu)}{f(x)}$$



$$\Rightarrow f(\mu|x) \propto \underbrace{f(x|\mu)}_{\text{Same as integrand of (1) a).}} \cdot f(\mu)$$

In this case $f(\mu)$ would play the role of the prior of μ .
 By the Algebra of (1) a), we know then that

$$f(\mu|x) \propto \exp \left\{ -\frac{1}{2} \cdot \left[\frac{\left(\mu - \left(\frac{Ax + Dm}{A+D} \right) \right)^2}{AD / (A+D)} \right] \right\},$$

a Normal $\left(\frac{Ax + Dm}{A+D} / \frac{AD}{(A+D)} \right)$ kernel

$$\mu \sim \text{N}(\bar{\mu}, \frac{1}{AD/(A+D)})$$

Thus, we know $\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \cdot \frac{\left(\mu - \left(\frac{Ax+Dm}{A+D}\right)\right)^2}{AD/(A+D)}\right\} = \sqrt{2\pi \left(\frac{AD}{A+D}\right)}$

Since $\int_{-\infty}^{\infty} f(\mu|x) \cdot d\mu = 1$, as it is a density

function, then:

$$f(\mu|x) = \frac{1}{\sqrt{2\pi \left(\frac{AD}{A+D}\right)}} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{\left(\mu - \left(\frac{Ax+Dm}{A+D}\right)\right)^2}{AD/(A+D)}\right\}$$

◀ ▶ ⌂ ⌂ ⌂ ⌂

② Our estimate of the variance would decrease from $\sigma_0^2 = \bar{p}(1-\bar{p})/90$ to $\sigma_1^2 = \bar{p}(1-\bar{p})/180$ which, once plugged in equation (7.20) would result in weaker shrinking of \hat{p}_i 's towards the mean. More explicitly:

$$\hat{p}_i^{\text{IS}} = \bar{p} + \left[1 - \frac{(N-3)}{\sum(p_i - \bar{p})^2} \cdot \sigma_0^2 \right] \cdot (p_i - \bar{p})$$

$$\text{If } \sigma_1^2 < \sigma_0^2 \Rightarrow \left[1 - \frac{(N-3)}{\sum(p_i - \bar{p})^2} \sigma_1^2 \right] > \left[1 - \frac{(N-3)}{\sum(p_i - \bar{p})^2} \sigma_0^2 \right]$$

⑤ ⑥

⑦ ⑧ ⑨

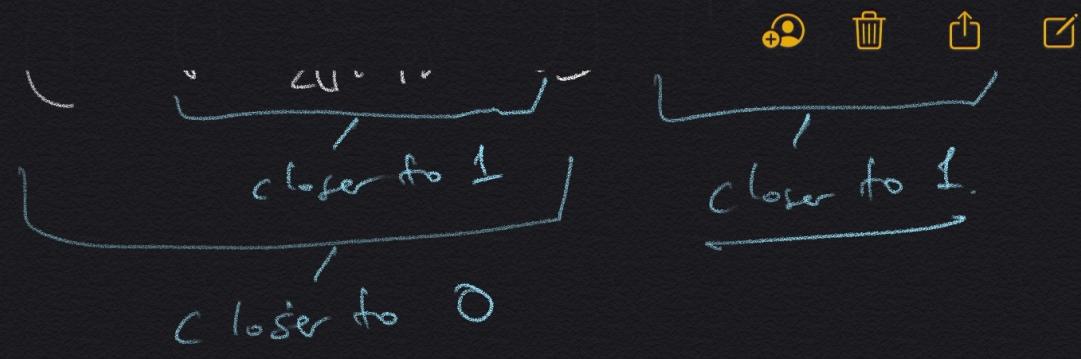
$$\hat{p}_i = \bar{p} + \frac{\sum (p_i - \bar{p})}{N}$$

$$\text{If } \sigma_i^2 < \sigma_0^2 \Rightarrow \left[1 - \frac{(N-3)}{\sum (p_i - \bar{p})^2} \sigma_i^2 \right] > \left[1 - \frac{(N-3)}{\sum (p_i - \bar{p})^2} \sigma_0^2 \right]$$

$$\text{Thus, } \hat{p}_i = \bar{p} \left[1 - \underbrace{\left(1 - \frac{(N-3)}{\sum (p_i - \bar{p})^2} \sigma_i^2 \right)}_{\text{closer to 1}} \right] + \underbrace{\left[1 - \frac{(N-3)}{\sum (p_i - \bar{p})^2} \cdot \sigma_0^2 \right]}_{\text{closer to 0}} \cdot p_i$$

\hat{p}_i is more similar to p_i and less similar to \bar{p} .

... the number of observations



\hat{p}_i is becomes more similar to p_i and less similar to \bar{P} .

You can substitute the number of observations (from 90 to 180) in the Jupyter notebook from Lecture 15 to verify this.

Since we calculated using 7.20 for
lecture 15, I created a New Jupyter notebook
which gets the SS estimator using 7.23, as
performed on table 7.1. By the way, there
is a typo on 7.23. Solving 7.21 for
 π_i and substituting x_i for \hat{w}_i^{SS} , we get
the correct version of 7.23. It properly reproduces
the values of table 7.1. More explicitly:

the correct version - 7.1.
the values of table 7.1. More explicitly:

$$\hat{\mu}_i^{JS} = 2(n+0.5)^{1/2} \cdot \sin^{-1} \left[\left(\frac{n \cdot \hat{p}_i^{JS} + 0.375}{n+0.75} \right)^{1/2} \right]$$

$$\frac{\hat{\mu}_i^{JS}}{2(n+0.5)^{1/2}} = \sin^{-1} \left[\left(\frac{n \cdot \hat{p}_i^{JS} + 0.375}{n+0.75} \right)^{1/2} \right]$$

$$\sin \left(\frac{\hat{\mu}_i^{JS}}{2(n+0.5)^{1/2}} \right) = \left(\frac{n \cdot \hat{p}_i^{JS} + 0.375}{n+0.75} \right)^{1/2}$$

$$\sin \left(\frac{\hat{\mu}_i^{JS}}{2(n+0.5)^{1/2}} \right)^2 = \frac{n \cdot \hat{p}_i^{JS} + 0.375}{n+0.75}$$

$\hat{\mu}_i^{JS}$

$$\sin\left(\frac{\hat{\mu}_i^{JS}}{2(n+0.5)^{1/2}}\right) = \sqrt{\frac{n \cdot \hat{p}_i^{JS} + 0.375}{n + 0.75}}$$

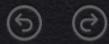
$$\sin^2\left(\frac{\hat{\mu}_i^{JS}}{2(n+0.5)^{1/2}}\right) = \frac{n \cdot \hat{p}_i^{JS} + 0.375}{n + 0.75}$$

$$\hat{p}_i^{JS} = \frac{1}{n} \left[(n+0.75) \cdot \sin^2\left(\frac{\hat{\mu}_i^{JS}}{2(n+0.5)^{1/2}}\right) - 0.375 \right].$$

④ See Jupyter Notebook.



⑤ I'd tell my brother-in-law that if the other players' batting average had been higher than that of player 4 (meaning pitchers are not very good this season), the JS estimator would not have let his player have an exceptionally low average. The cross-player information, leveraged by the JS estimator, gives us an idea of how good/poor pitchers vs. batters are in general.





⑥ Starting from (7.37) and (7.38)

$$\hat{\beta} \sim N_p(\beta, \sigma^2 \cdot S^{-1}) \text{ and } \beta \sim N_p(0, \frac{\sigma^2}{\lambda} \cdot I)$$

$$\Rightarrow f(\beta | \hat{\beta}) \propto f(\hat{\beta} | \beta) \cdot f(\beta)$$

$$f(\beta | \hat{\beta}) \propto \exp \left\{ -\frac{1}{2} \cdot \left[(\beta - \hat{\beta})^T (\sigma^2 S^{-1})^{-1} (\beta - \hat{\beta}) + \beta^T \left(\frac{\sigma^2}{\lambda} \cdot I \right)^{-1} \beta \right] \right\}$$

$$f(\beta | \hat{\beta}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \left[(\beta - \hat{\beta})^T S \cdot (\beta - \hat{\beta}) + \beta^T \cdot \lambda I \cdot \beta \right] \right\}$$

$\Gamma_n \Gamma_D \dots \Gamma_{n-1} \Gamma_{n-2} \dots \Gamma_1 \hat{\beta} \}$



$$f(\beta | \hat{\beta}) \propto \exp(-\frac{1}{2\sigma^2} [$$

$$\alpha \exp\left\{-\frac{1}{2\sigma^2} \left[\hat{\beta}^T (S + \lambda I) \hat{\beta} - \hat{\beta}^T S \beta - \beta^T S \hat{\beta} \right] \right\}$$

Since we're working with expressions that are proportional to $f(\beta | \hat{\beta})$, we can multiply (add to the argument inside of $\exp\{\cdot\}$) pretty much any expression that does not depend on β . Thus, completing a square ...

$$f(\beta | \hat{\beta}) \propto \exp\left\{-\frac{1}{2\sigma^2} \left[\hat{\beta}^T (S + \lambda I) \hat{\beta} - \hat{\beta}^T S \beta - \beta^T S \hat{\beta} + \hat{\beta}^T S \cdot (S + \lambda I)^{-1} S \hat{\beta} \right] \right\}$$



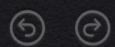
comp. ...

$$f(\beta | \hat{\beta}) \propto \exp \left\{ -\frac{1}{2\sigma^2} \cdot \left[\beta^T (S + \lambda I) \beta - \hat{\beta}^T S \beta - \beta^T S \hat{\beta} + \hat{\beta}^T S \cdot (S + \lambda I)^{-1} S \hat{\beta} \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2\sigma^2} \cdot \left[\beta^T (S + \lambda I) \beta - \hat{\beta}^T S \cdot (S + \lambda I)^{-1} (S + \lambda I) \beta - \beta^T (S + \lambda I) (S + \lambda I)^{-1} S \hat{\beta} \right. \right. \\ \left. \left. + \hat{\beta}^T \cdot S \cdot (S + \lambda I)^{-1} (S + \lambda I) \cdot (S + \lambda I)^{-1} S \hat{\beta} \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2\sigma^2} \cdot \left[\left(\beta^T - \hat{\beta}^T S \cdot (S + \lambda I)^{-1} \right) (S + \lambda I) \beta \right. \right. \\ \left. \left. - \left(\beta^T - \hat{\beta}^T S \cdot (S + \lambda I)^{-1} \right) (S + \lambda I) (S + \lambda I)^{-1} S \hat{\beta} \right] \right\}$$

$$\propto \exp \left\{ -\frac{1}{2\sigma^2} \cdot \left[\left(\beta^T - \hat{\beta}^T S \cdot (S + \lambda I)^{-1} \right) (S + \lambda I) \left(\beta - (S + \lambda I)^{-1} S \hat{\beta} \right) \right] \right\}$$



$$\propto \exp\left(-\frac{1}{2\sigma^2} \cdot \left[(\beta^T - \hat{\beta}^T S (S + \lambda I)^{-1}) (S + \lambda I) (\beta - (S + \lambda I)^{-1} S \hat{\beta}) \right]\right)$$

Recalling that $S = X^T X$ is symmetric (and since λI is diagonal), $S + \lambda I$ is also symmetric. Thus $S^T = S$ and $(S + \lambda I)^T = (S + \lambda I)$

$$\propto \exp\left\{-\frac{1}{2\sigma^2} \left[(\beta - (S + \lambda I)^{-1} S \hat{\beta})^T (S + \lambda I) (\beta - (S + \lambda I)^{-1} S \hat{\beta}) \right]\right\}$$

Which is a $N_p((S + \lambda I)^{-1} S \hat{\beta}, \sigma^2 \cdot (S + \lambda I)^{-1})$ kernel.

$$\text{Thus, } E(\beta | \hat{\beta}) = (S + \lambda I)^{-1} S \hat{\beta}$$



- 7) a) → Check all calculations on Ipyter Notebook
b) → for Lecture 16, part 1.
- 8) Explained on Note 7 of the notes &
details section of Chapter 7.
- 9) We did this on Lecture 16.
Check Lecture 16's notes.

