14 November 2019, 8:15

Chapter 7 solutions

① $\mu \sim N(M, A)$

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

a) What is the marginal distribution of x?

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

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

8:22 Thu 14 Nov 44%

use conditional probability to -

$$f(X, \mu) = f(x|\mu) \cdot f(\mu)$$

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

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

Thus, using the Normal pdf.

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

... write the integrand

$$f(x) = \int_{-\infty} \frac{}{\sqrt{2\pi D}} \cdots$$

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

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\cdot\frac{(\mu - E(\mu))^2}{V(\mu)}\right\}\cdot d\mu = \frac{1}{\sqrt{2\pi\cdot V(\mu)}}$$

. Ok, let's start...

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

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

$-\infty$

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

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

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

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

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

$$2\pi\sqrt{AD}\ {}_{-\infty}$$

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

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

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

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

$$= \frac{1}{\cdots} \cdot \int^{\infty} \exp\left\{-\frac{1}{2}\cdot\left[\frac{\left(\mu - \left(\frac{Ax+Dm}{(A+D)}\right)\right)^2}{AD/(A+D)} + \frac{\cancel{A^2x^2} + A\cancel{D}m^2 + \cancel{A}Dx^2 + \cancel{D^2m^2} - \cancel{A^2x^2} - 2\cancel{A}x\cancel{D}m - \cancel{D^2m^2}}{\cancel{AD}}\right]\right\} d\mu$$

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

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

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

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

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

$$\sqrt{2\pi}\sqrt{AD} \quad -\infty \qquad (AD/(A+D)$$

$$= \frac{\exp\left\{-\frac{1}{2}\cdot\left[(x-m)^2\right]\right\}}{2\pi\sqrt{AD}} \cdot \underbrace{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\cdot\left[\frac{\left(\mu-\left(\frac{Ax+Dm}{(A+D)}\right)\right)^2}{AD/(A+D)}\right]\right\}}_{N\left(\frac{Ax+Dm}{A+D},\ \frac{AD}{A+D}\right)\ \text{kernel.}} \cdot d\mu$$

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

$$= \frac{\exp\left\{-\frac{1}{2}\cdot\left[(x-m)^2\right]\right\}}{\ldots}$$

8:23 Thu 14 Nov 44%

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

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

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

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

$$\Rightarrow \quad f(\mu \mid x) \propto \underbrace{f(x \mid \mu) \cdot f(\mu)}$$

Same as integrand of ① a).

In this case $f(\mu)$ would play the role of the prior of μ.

By the Algebra of ① a), we know then that

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

a $\text{Normal}\left(\dfrac{Ax + Dm}{A+D}, \dfrac{AD}{(A+D)} \right)$ kernel

$\left(A+D \;/\; (A+D) \right)$

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

- Since $\displaystyle\int_{-\infty}^{\infty} f(\mu|x)\cdot d\mu = 1$, as it is a <u>density function</u>, then:

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

(2) Our estimate of the variance would

<u>decrease</u> from $\bar{\sigma}_0^2 = \bar{p}(1-\bar{p})/90$ to $\sigma_i^2 = \bar{p}(1-p)/180$

which, once plugged in equation (7.20) would result

in <u>weaker</u> shrinking of $\hat{p}_i^{JS}$ towards the mean.

More explicitly:

$$\hat{p}_i^{JS} = \bar{p} + \left[1 - \frac{(N-3)}{\Sigma(p_i-\bar{p})^2} \cdot \bar{\sigma}_0^2\right] \cdot (p_i - \bar{p})$$

If $\sigma_i^2 < \bar{\sigma}_0^2 \Rightarrow \left[1 - \frac{(N-3)}{\Sigma(p_i-\bar{p})^2} \sigma_i^2\right] > \left[1 - \frac{(N-3)}{\Sigma(p_i-\bar{p})^2} \bar{\sigma}_0^2\right]$

$\hat{p}_i \ldots = [\ldots (\ldots \Sigma(p_i-\bar{p})^2 \ldots)]$

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

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

$\hat{p}_i^{JS}$ becomes <u>more similar</u> to p_i and <u>less similar</u> to $\bar{p}$.

... the number of observations

closer to 1 closer to 1.

closer to 0

$\hat{p}_i$ 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.

3 Since we calculated using 7.20 for Lecture 15, I created a <u>New Jupyter Notebook</u> which gets the JS estimator using 7.23, as performed on table <u>7.1</u>. By the way, there is a <u>typo</u> on 7.23. Solving 7.21 for μ_i and substituting x_i for $\hat{\mu}_i^{JS}$, we get the correct version of <u>7.23</u>. It properly reproduces the values of table <u>7.1</u>. More explicitly:

8:24 Thu 14 Nov 43%

the correct version ...

the values of table 7.1. More explicitly:

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

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

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

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

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

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

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

(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)

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

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

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

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

$$f(\beta|\hat\beta) \propto \exp\left(-\frac{1}{2\sigma^2}\left[\quad \cdots \quad \right]\right)$$

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

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...

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

comp[illegible]

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

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

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

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

8:25 Thu 14 Nov 42%

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

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

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

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

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

8:25 Thu 14 Nov 42%

⑦ a) → Check all calculations on Jupyter 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.