Chapter 1 Introduction

Yue Yu

1.2 Probability Theory

(1.39)

$$\begin{aligned}
\mathrm{var}[f] &= \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2] \\
&= \mathbb{E}[f(x)^2 - 2\mathbb{E}[f(x)]f(x) + \mathbb{E}[f(x)]^2] \\
&= \mathbb{E}[f(x)^2] - \mathbb{E}[2\mathbb{E}[f(x)]f(x)] + \mathbb{E}[\mathbb{E}[f(x)]^2] \\
&= \mathbb{E}[f(x)^2] - 2\mathbb{E}[f(x)]^2 + \mathbb{E}[f(x)]^2 \\
&= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2.
\end{aligned}$$

(1.48)

Denote

$$I = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2} x^2 \right) \mathrm{d}x, \qquad (*)$$

then

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2} x^2 - \frac{1}{2\sigma^2} y^2 \right) \mathrm{d}x\, \mathrm{d}y.$$

Applying the transformation $x = r\cos\theta$, $y = r\sin\theta$, and making use of (1.27), we have

$$\begin{aligned}
I^2 &= \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{1}{2\sigma^2}(r\cos\theta)^2 - \frac{1}{2\sigma^2}(r\sin\theta)^2 \right) \left| \frac{\partial(x,y)}{\partial(r,\theta)} \right| \mathrm{d}r\, \mathrm{d}\theta \\
&= \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{r^2}{2\sigma^2} \right) r\, \mathrm{d}r\, \mathrm{d}\theta.
\end{aligned}$$

Then, by substituting $u = r^2$, we obtain

$$\begin{aligned}
I^2 &= \int_0^{2\pi} \int_0^{\infty} \exp\left(-\frac{u}{2\sigma^2} \right) u^{1/2} \left| \frac{\mathrm{d}r}{\mathrm{d}u} \right| \mathrm{d}u\, \mathrm{d}\theta \\
&= \int_0^{2\pi} \int_0^{\infty} \frac{1}{2} \exp\left(-\frac{u}{2\sigma^2} \right) \mathrm{d}u\, \mathrm{d}\theta \\
&= -\sigma^2 \exp\left(-\frac{u}{2\sigma^2} \right) \Bigg|_0^{\infty} \int_0^{2\pi} \mathrm{d}\theta \\
&= 2\pi\sigma^2,
\end{aligned}$$

which, together with (*), implies that

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2} x^2 \right) \mathrm{d}x = I = (2\pi\sigma^2)^{1/2}.$$

Finally, by further substituting $x = z - \mu$, we have

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{(z-\mu)^2}{2\sigma^2} \right) \mathrm{d}z \\
&= \int_{-\infty}^{\infty} \mathcal{N}(z|\mu, \sigma^2)\, \mathrm{d}z.
\end{aligned}
$$

(1.49)

$$
\begin{aligned}
\mathbb{E}[x] &= \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x \,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right) x \,\mathrm{d}x.
\end{aligned}
$$

Applying the transformation $z = x - \mu$, we obtain

$$
\begin{aligned}
\mathbb{E}[x] &= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{z^2}{2\sigma^2} \right) (z + \mu) \,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{z^2}{2\sigma^2} \right) z \,\mathrm{d}z + \mu \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{z^2}{2\sigma^2} \right) \mathrm{d}z \\
&= \mu
\end{aligned}
$$

where in the second last step, the first term vanishes because it integrates over an odd function.

(1.50)

Recall

$$
\int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)\, \mathrm{d}x = 1.
$$

Differentiating both sides with respect to σ^2, we have

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \sigma^2} \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2)\, \mathrm{d}x \\
&= - \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right) \frac{(\sigma^2)^{-1}}{2} \,\mathrm{d}x + \int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right) \frac{(x-\mu)^2 (\sigma^2)^{-2}}{2} \,\mathrm{d}x \\
&= - \frac{1}{2} \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) (\sigma^2)^{-1} \,\mathrm{d}x + \frac{1}{2} \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) (x-\mu)^2 (\sigma^2)^{-2} \,\mathrm{d}x.
\end{aligned}
$$

Multiplying $2\sigma^4$ on both sides, we obtain

$$
\begin{aligned}
0 &= - \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) \sigma^2 \,\mathrm{d}x + \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) (x-\mu)^2 \,\mathrm{d}x \\
&= -\sigma^2 \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) \,\mathrm{d}x + \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x^2 \,\mathrm{d}x - 2\mu \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) x \,\mathrm{d}x + \mu^2 \int_{-\infty}^{\infty} \mathcal{N}(x|\mu, \sigma^2) \,\mathrm{d}x \\
&= -\sigma^2 + \mathbb{E}[x^2] - 2\mu\mathbb{E}[x] + \mu^2 \\
&= -\sigma^2 + \mathbb{E}[x^2] - \mu^2.
\end{aligned}
$$

Therefore, $\mathbb{E}[x^2] = \mu^2 + \sigma^2$.

(1.51)

Combining (1.49) and (1.50), it is easy to see that

$$
\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2.
$$

(1.55)

Recall the log likelihood function

$$\ell(\mu, \sigma) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi).$$

Solving for μ_{ML}, we obtain

$$\begin{aligned}
\mu_{\mathrm{ML}} &= \arg_{\mu}\left(\frac{\partial}{\partial\mu}\ell(\mu,\sigma) = 0\right)\Bigg|_{\sigma^2=\sigma^2_{\mathrm{ML}}} \\
&= \arg_{\mu}\left(\frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n - \mu) = 0\right) \\
&= \frac{1}{N}\sum_{n=1}^{N} x_n.
\end{aligned}$$

(1.56)

$$\begin{aligned}
\sigma^2_{\mathrm{ML}} &= \arg_{\sigma^2}\left(\frac{\partial}{\partial\sigma^2}\ell(\mu,\sigma) = 0\right)\Bigg|_{\mu=\mu_{\mathrm{ML}}} \\
&= \arg_{\sigma^2}\left(\frac{1}{2}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2(\sigma^2)^{-2} - \frac{N}{2}\frac{1}{\sigma^2} = 0\right) \\
&= \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{\mathrm{ML}})^2.
\end{aligned}$$

(1.57)

$$\begin{aligned}
\mathbb{E}[\mu_{\mathrm{ML}}] &= \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} x_n\right] \\
&= \frac{1}{N}\mathbb{E}\left[\sum_{n=1}^{N} x_n\right] \\
&= \frac{1}{N}N\mu \\
&= \mu
\end{aligned}$$

where the third step took advantage of the fact that the data points are independent, identically distributed, so that $\mathbb{E}[x_n] = \mu$ for any $n \in \{1, \ldots, N\}$.

(1.58)

$$
\begin{aligned}
\mathbb{E}[\sigma^2_{\mathrm{ML}}] &= \mathbb{E}\bigg[\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{\mathrm{ML}})^2\bigg] \\
&= \mathbb{E}\bigg[\frac{1}{N}\sum_{n=1}^{N}(x_n-\frac{1}{N}\sum_{m=1}^{N}x_m)^2\bigg] \\
&= \frac{1}{N}\mathbb{E}\bigg[\sum_{n=1}^{N}x_n^2-\frac{2}{N}\bigg(\sum_{n=1}^{N}x_n\bigg)^2+\frac{1}{N^2}\sum_{n=1}^{N}\bigg(\sum_{m=1}^{N}x_m\bigg)^2\bigg] \\
&= \frac{1}{N}\mathbb{E}\bigg[\sum_{n=1}^{N}x_n^2\bigg]-\frac{2}{N^2}\mathbb{E}\bigg[\bigg(\sum_{n=1}^{N}x_n\bigg)^2\bigg]+\frac{1}{N^3}\mathbb{E}\bigg[\sum_{n=1}^{N}\bigg(\sum_{m=1}^{N}x_m\bigg)^2\bigg] \\
&= \frac{1}{N}\mathbb{E}\bigg[\sum_{n=1}^{N}x_n^2\bigg]-\frac{2}{N^2}\mathbb{E}\bigg[\bigg(\sum_{n=1}^{N}x_n\bigg)^2\bigg]+\frac{1}{N^2}\mathbb{E}\bigg[\bigg(\sum_{m=1}^{N}x_m\bigg)^2\bigg] \\
&= \frac{1}{N}\mathbb{E}\bigg[\sum_{n=1}^{N}x_n^2\bigg]-\frac{1}{N^2}\mathbb{E}\bigg[\bigg(\sum_{n=1}^{N}x_n\bigg)^2\bigg] \\
&= \frac{1}{N}N(\mu^2+\sigma^2)-\frac{1}{N^2}(N^2\mu^2+N\sigma^2) \\
&= \frac{N-1}{N}\sigma^2.
\end{aligned}
$$

In the second last step, for the first term, analogous to the proof of (1.57), we made use of the i.i.d. assumption such that $\mathbb{E}[x_n^2]=\mu^2$ for any $n\in\{1,\ldots,N\}$. In terms of the second term, for any $m,n\in\{1,\ldots,N\}$, we have $\mathbb{E}[x_nx_m]=\mu^2+I_{nm}\sigma^2$, where $I_{nm}=1$ if $n=m$, and $I_{nm}=0$ otherwise. If we expand the square of sum, there are N out of N^2 terms with $n=m$, and N^2-N terms with $n\neq m$, which gives rise to the result.

1.5 Decision Theory

(1.88)

Let

$$G(y,\mathbf{x})=\int\{y(\mathbf{x})-t\}^2p(\mathbf{x},t)\,\mathrm{d}t$$

such that

$$\mathbb{E}[L]=\int G(y,\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Using Euler-Lagrange equation, we obtain

$$\frac{\partial G(y,\mathbf{x})}{\partial y(\mathbf{x})}=2\int\{y(\mathbf{x})-t\}p(\mathbf{x},t)\,\mathrm{d}t=0.$$

(1.90)

$$
\begin{aligned}
\mathbb{E}[L] &= \iint \{y(\mathbf{x}) - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \iint \{y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}] + \mathbb{E}_t[t|\mathbf{x}] - t\}^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \iint \left\{(y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])^2 + 2(y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])(\mathbb{E}_t[t|\mathbf{x}] - t) + (\mathbb{E}_t[t|\mathbf{x}] - t)^2\right\} p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \iint (y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \iint (\mathbb{E}_t[t|\mathbf{x}] - t)^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \\
&\quad 2\iint (y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])(\mathbb{E}_t[t|\mathbf{x}] - t) p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t.
\end{aligned}
$$

Expanding the third term while omitting the coefficient 2, we have

$$
\begin{aligned}
&\iint (y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])(\mathbb{E}_t[t|\mathbf{x}] - t) p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \iint y(\mathbf{x})\mathbb{E}_t[t|\mathbf{x}] p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t - \iint y(\mathbf{x}) t p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t - \iint \mathbb{E}_t[t|\mathbf{x}]^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \iint \mathbb{E}_t[t|\mathbf{x}] t p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \mathbb{E}_t[t|\mathbf{x}] \int y(\mathbf{x}) \int p(\mathbf{x}, t)\,\mathrm{d}t\,\mathrm{d}\mathbf{x} - \int y(\mathbf{x}) p(\mathbf{x}) \int t p(t|\mathbf{x})\,\mathrm{d}t\,\mathrm{d}\mathbf{x} - \mathbb{E}_t[t|\mathbf{x}]^2 + \mathbb{E}_t[t|\mathbf{x}] \int t p(t|\mathbf{x}) \int p(\mathbf{x})\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \mathbb{E}_t[t|\mathbf{x}]\mathbb{E}_y[\mathbf{x}] - \mathbb{E}_t[t|\mathbf{x}]\mathbb{E}_y[\mathbf{x}] - \mathbb{E}_t[t|\mathbf{x}]^2 + \mathbb{E}_t[t|\mathbf{x}]^2 \\
&= 0.
\end{aligned}
$$

Thus, the cross term vanishes and we obtain

$$
\begin{aligned}
\mathbb{E}[L] &= \iint (y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \iint (\mathbb{E}_t[t|\mathbf{x}] - t)^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \int (y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])^2 \int p(t|\mathbf{x})\,\mathrm{d}t\,\mathrm{d}\mathbf{x} + \iint (\mathbb{E}_t[t|\mathbf{x}] - t)^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \int (y(\mathbf{x}) - \mathbb{E}_t[t|\mathbf{x}])^2\,\mathrm{d}\mathbf{x} + \iint (\mathbb{E}_t[t|\mathbf{x}] - t)^2 p(\mathbf{x}, t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t.
\end{aligned}
$$

Note that this result is different from the one in the book.