Model Answers to Exercise 1 of February 23, 2017

http://www.isi.ee.ethz.ch/teaching/courses/it2.html

Problem 1 ***Differential Entropy***

a) Exponential distribution:

$$
\begin{aligned}
h(X) &= -\int_0^\infty \lambda e^{-\lambda x} \log\bigl(\lambda e^{-\lambda x}\bigr)\,\mathrm{d}x \\
&= -\log\lambda \cdot \underbrace{\int_0^\infty \lambda e^{-\lambda x}\,\mathrm{d}x}_{=1} - \int_0^\infty \lambda e^{-\lambda x}(-\lambda x)\log e\,\mathrm{d}x \\
&= -\log\lambda + \lambda^2\log e \cdot \int_0^\infty x e^{-\lambda x}\,\mathrm{d}x \\
&\overset{\text{(i)}}{=} -\log\lambda + \lambda^2 \log e \cdot \underbrace{x\cdot\frac{e^{-\lambda x}}{-\lambda}\bigg|_{x=0}^{\infty}}_{=0} - \lambda^2\log e\cdot\int_0^\infty \frac{e^{-\lambda x}}{-\lambda}\,\mathrm{d}x \\
&= -\log\lambda + \log e \cdot \underbrace{\int_0^\infty \lambda e^{-\lambda x}\,\mathrm{d}x}_{=1} \\
&= \log\frac{e}{\lambda},
\end{aligned}
$$

where (i) follows from integration by parts.

b) Laplace distribution:

$$
\begin{aligned}
h(X) &= -\int_{-\infty}^\infty \frac{\lambda}{2}e^{-\lambda|x|}\log\left(\frac{\lambda}{2}e^{-\lambda|x|}\right)\mathrm{d}x \\
&\overset{\text{(i)}}{=} -2\cdot\int_0^\infty \frac{\lambda}{2}e^{-\lambda x}\log\left(\frac{\lambda}{2}e^{-\lambda x}\right)\mathrm{d}x \\
&= -\int_0^\infty \lambda e^{-\lambda x}\log\left(\frac{\lambda}{2}e^{-\lambda x}\right)\mathrm{d}x \\
&= -\log\frac{1}{2}\cdot\underbrace{\int_0^\infty \lambda e^{-\lambda x}\,\mathrm{d}x}_{=1} - \int_0^\infty \lambda e^{-\lambda x}\log\bigl(\lambda e^{-\lambda x}\bigr)\,\mathrm{d}x \\
&\overset{\text{(ii)}}{=} \log 2 + \log\frac{e}{\lambda} \\
&= \log\frac{2e}{\lambda},
\end{aligned}
$$

where (i) holds because the integrand is an even function; and (ii) follows from Part a).

c) The sum of two independent Gaussian random variables is also Gaussian. Since X_1 and X_2 are independent, the variance of $X = X_1 + X_2$ is $\sigma_1^2 + \sigma_2^2$, and we obtain

$$h(X) = \frac{1}{2}\log\bigl(2\pi e(\sigma_1^2 + \sigma_2^2)\bigr).$$

Problem 2 ***Scaling Property***

Let $\mathbf{Y} = \mathsf{A}\mathbf{X}$. The density of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det \mathsf{A}|}\, f_{\mathbf{X}}(\mathsf{A}^{-1}\mathbf{y}),$$

and hence

$$\begin{aligned}
h(\mathsf{A}\mathbf{X}) &= -\int_{\mathbb{R}^n} f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}}(\mathbf{y})\,\mathrm{d}\mathbf{y} \\
&= -\int_{\mathbb{R}^n} \frac{1}{|\det \mathsf{A}|} f_{\mathbf{X}}(\mathsf{A}^{-1}\mathbf{y}) \Bigl(\log f_{\mathbf{X}}(\mathsf{A}^{-1}\mathbf{y}) - \log|\det \mathsf{A}| \Bigr)\,\mathrm{d}\mathbf{y} \\
&\overset{\text{(i)}}{=} -\int_{\mathbb{R}^n} \frac{1}{|\det \mathsf{A}|} f_{\mathbf{X}}(\mathbf{x}) \Bigl(\log f_{\mathbf{X}}(\mathbf{x}) - \log|\det \mathsf{A}| \Bigr) |\det \mathsf{A}|\,\mathrm{d}\mathbf{x} \\
&= -\int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) \log f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \log|\det \mathsf{A}| \cdot \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= h(\mathbf{X}) + \log|\det \mathsf{A}|,
\end{aligned}$$

where (i) follows from integration by substitution.

Problem 3 ***AEP for Gaussian Random Variables***

a) Since $X_1, \ldots, X_n$ are IID Gaussian, we have

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x_i-\mu)^2}{2\sigma^2}},$$

and

$$\begin{aligned}
Y &= -\frac{1}{n}\log \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} - \frac{1}{2}\log\bigl(2\pi e\sigma^2\bigr) \\
&= -\frac{1}{n}\sum_{i=1}^n \log\biggl(\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \biggr) - \frac{1}{2}\log\bigl(2\pi e\sigma^2\bigr) \\
&= \frac{1}{2}\log\bigl(2\pi\sigma^2\bigr) - \frac{1}{n}\sum_{i=1}^n \frac{-(X_i-\mu)^2}{2\sigma^2}\log e - \frac{1}{2}\log\bigl(2\pi e\sigma^2\bigr) \\
&= -\frac{1}{2}\log e + \frac{\log e}{2n} \cdot \sum_{i=1}^n \biggl(\frac{X_i-\mu}{\sigma}\biggr)^2 \\
&= \frac{\log e}{2n} \cdot \Biggl(\sum_{i=1}^n \biggl(\frac{X_i-\mu}{\sigma}\biggr)^2 - n \Biggr).
\end{aligned}$$

Introducing $Z_i = \frac{X_i-\mu}{\sigma}$ leads to

$$Y = \frac{\log e}{2n} \cdot \Biggl(\sum_{i=1}^n Z_i^2 - n \Biggr).$$

Note that the Z_i are IID $\mathcal{N}(0,1)$ because the X_i are IID $\mathcal{N}(\mu,\sigma^2)$. Thus, $\chi_n^2 = \sum_{i=1}^n Z_i^2$ is central chi-square distributed with n degrees of freedom and has probability density function

$$f_{\chi_n^2}(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\bigl(\frac{n}{2}\bigr)}\, x^{\frac{n}{2}-1} e^{-\frac{x}{2}}\, \mathrm{I}\{x>0\}.$$

Since

$$Y = \frac{\log e}{2n}\cdot\bigl(\chi_n^2 - n\bigr),$$

we finally obtain

$$\begin{aligned} f_Y(y) &= \frac{2n}{\log e}\cdot f_{\chi_n^2}\biggl(\frac{2ny}{\log e}+n\biggr) \\ &= \frac{2n}{\log e\cdot 2^{\frac{n}{2}}\Gamma\bigl(\frac{n}{2}\bigr)}\biggl(\frac{2ny}{\log e}+n\biggr)^{\frac{n}{2}-1} e^{-\frac{ny}{\log e}-\frac{n}{2}}\, \mathrm{I}\biggl\{\frac{2ny}{\log e}+n>0\biggr\}, \end{aligned}$$

which does not depend on μ or σ^2.

b) We have

$$\begin{aligned} Y &= -\frac{1}{n}\log\Biggl(\frac{1}{\sqrt{(2\pi)^n\det\mathsf{K}}}\, e^{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^\mathsf{T}\mathsf{K}^{-1}(\mathbf{X}-\boldsymbol{\mu})}\Biggr) - \frac{1}{n}\cdot\frac{1}{2}\log\bigl((2\pi e)^n\det\mathsf{K}\bigr) \\ &= \frac{1}{2}\log(2\pi) + \frac{1}{2n}\log(\det\mathsf{K}) + \frac{1}{2n}(\mathbf{X}-\boldsymbol{\mu})^\mathsf{T}\mathsf{K}^{-1}(\mathbf{X}-\boldsymbol{\mu})\log e - \frac{1}{2}\log(2\pi e) - \frac{1}{2n}\log(\det\mathsf{K}) \\ &= -\frac{1}{2}\log e + \frac{\log e}{2n}(\mathbf{X}-\boldsymbol{\mu})^\mathsf{T}\mathsf{K}^{-1}(\mathbf{X}-\boldsymbol{\mu}). \end{aligned}$$

Because K is positive definite, there exists a $n\times n$ matrix A such that $\mathsf{K}=\mathsf{A}\mathsf{A}^\mathsf{T}$. Consequently, $\mathsf{K}^{-1}=\mathsf{A}^{-\mathsf{T}}\mathsf{A}^{-1}$. Introducing $\mathbf{Z}=\mathsf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu})$ leads to

$$\begin{aligned} Y &= -\frac{1}{2}\log e + \frac{\log e}{2n}(\mathbf{X}-\boldsymbol{\mu})^\mathsf{T}\mathsf{A}^{-\mathsf{T}}\mathsf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \\ &= -\frac{1}{2}\log e + \frac{\log e}{2n}\cdot\mathbf{Z}^\mathsf{T}\mathbf{Z} \\ &= \frac{\log e}{2n}\cdot\Biggl(\sum_{i=1}^n Z_i^2 - n\Biggr). \end{aligned}$$

Note that $\mathbf{Z}$ is Gaussian because it is an affine transformation of a Gaussian random vector, so the distribution of $\mathbf{Z}$ is determined by its mean vector and its covariance matrix. We have

$$\mathsf{E}[\mathbf{Z}] = \mathsf{E}\bigl[\mathsf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu})\bigr] = \mathsf{A}^{-1}\,\mathsf{E}[\mathbf{X}-\boldsymbol{\mu}] = \mathsf{A}^{-1}\cdot\mathbf{0} = \mathbf{0}$$

and

$$\begin{aligned} \mathsf{K}_{\mathbf{ZZ}} &= \mathsf{E}[(\mathbf{Z}-\mathbf{0})(\mathbf{Z}-\mathbf{0})^\mathsf{T}] \\ &= \mathsf{E}[\mathbf{Z}\mathbf{Z}^\mathsf{T}] \\ &= \mathsf{E}\bigl[\mathsf{A}^{-1}(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\mathsf{T}\mathsf{A}^{-\mathsf{T}}\bigr] \\ &= \mathsf{A}^{-1}\,\mathsf{E}[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\mathsf{T}]\mathsf{A}^{-\mathsf{T}} \\ &= \mathsf{A}^{-1}\mathsf{K}\mathsf{A}^{-\mathsf{T}} \\ &= \mathsf{A}^{-1}\mathsf{A}\mathsf{A}^\mathsf{T}\mathsf{A}^{-\mathsf{T}} \\ &= \mathsf{I}. \end{aligned}$$

Therefore, the Z_i are IID $\mathcal{N}(0,1)$ and we get the same result as in Part a).

c) The density of $\mathbf{Y} = \mathsf{A}\mathbf{X}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\det \mathsf{A}|}\, f_{\mathbf{X}}(\mathsf{A}^{-1}\mathbf{y}),$$

and hence

$$\begin{aligned}
-\log f_{\mathbf{Y}}(\mathbf{Y}) - h(\mathbf{Y}) &= -\log\left(\frac{1}{|\det \mathsf{A}|}\, f_{\mathbf{X}}(\mathsf{A}^{-1}\mathbf{Y})\right) - h(\mathbf{Y}) \\
&= \log|\det \mathsf{A}| - \log f_{\mathbf{X}}(\mathbf{X}) - h(\mathsf{A}\mathbf{X}) \\
&\overset{\text{(i)}}{=} \log|\det \mathsf{A}| - \log f_{\mathbf{X}}(\mathbf{X}) - h(\mathbf{X}) - \log|\det \mathsf{A}| \\
&= -\log f_{\mathbf{X}}(\mathbf{X}) - h(\mathbf{X}),
\end{aligned}$$

where (i) follows from Problem 2.