Model Answers to Exercise 2 of March 2, 2017

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

Problem 1 — *Log-Concavity of Determinants*

Note that the inequality holds if K_1 or K_2 is singular: If $\lambda = 0$ or $\lambda = 1$, the claim is trivially true; if $\lambda \in (0,1)$, we have

$$\det\bigl(\lambda \mathsf{K}_1 + \bar{\lambda}\mathsf{K}_2\bigr) \stackrel{\text{(i)}}{\geq} 0 \stackrel{\text{(ii)}}{=} (\det \mathsf{K}_1)^{\lambda} (\det \mathsf{K}_2)^{\bar{\lambda}},$$

where (i) is true because the convex combination of two positive semidefinite matrices is again positive semidefinite, hence its determinant is nonnegative; and (ii) follows from the assumption that K_1 or K_2 is singular and from $\lambda \in (0,1)$.
From now on, we assume that both K_1 and K_2 are nonsingular. Let $\mathbf{X}_1 \sim \mathcal{N}(\mathbf{0}, \mathsf{K}_1)$, $\mathbf{X}_2 \sim \mathcal{N}(\mathbf{0}, \mathsf{K}_2)$, and $\theta \sim \text{Bernoulli}(\lambda)$ be independent. Because K_1 and K_2 are nonsingular, K_1 and K_2 are positive definite, so $\mathbf{X}_1$ and $\mathbf{X}_2$ have a probability density. Let $\mathbf{Z} = \mathbf{X}_\theta$, i.e.,

$$\mathbf{Z} = \begin{cases} \mathbf{X}_1 & \text{if } \theta = 1, \\ \mathbf{X}_2 & \text{if } \theta = 2. \end{cases}$$

The mean of $\mathbf{Z}$ is $\mathbf{0}$ since

$$\mathsf{E}[\mathbf{Z}] = \Pr[\theta = 1] \cdot \mathsf{E}[\mathbf{Z}|\theta = 1] + \Pr[\theta = 2] \cdot \mathsf{E}[\mathbf{Z}|\theta = 2] = \lambda \cdot \mathsf{E}[\mathbf{X}_1] + \bar{\lambda} \cdot \mathsf{E}[\mathbf{X}_2] = \mathbf{0}.$$

The covariance matrix of $\mathbf{Z}$ is

$$\begin{aligned} \mathsf{K}_{\mathbf{ZZ}} &= \mathsf{E}\bigl[(\mathbf{Z} - \mathbf{0})(\mathbf{Z} - \mathbf{0})^{\mathsf{T}}\bigr] \\ &= \mathsf{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}] \\ &= \Pr[\theta = 1] \cdot \mathsf{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\theta = 1] + \Pr[\theta = 2] \cdot \mathsf{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\theta = 2] \\ &= \lambda \cdot \mathsf{E}[\mathbf{X}_1\mathbf{X}_1^{\mathsf{T}}] + \bar{\lambda} \cdot \mathsf{E}[\mathbf{X}_2\mathbf{X}_2^{\mathsf{T}}] \\ &= \lambda \mathsf{K}_1 + \bar{\lambda}\mathsf{K}_2, \end{aligned}$$

which is positive definite because the convex combination of two positive definite matrices is again positive definite. Note that $\mathbf{Z}$ is in general not a Gaussian random vector. But since a Gaussian random vector maximizes differential entropy among all random vectors having the same covariance matrix, we have

$$h(\mathbf{Z}) \leq \frac{1}{2}\log\Bigl((2\pi e)^n \det(\lambda \mathsf{K}_1 + \bar{\lambda}\mathsf{K}_2)\Bigr).$$

Since conditioning does not increase entropy, we also have

$$h(\mathbf{Z}) \geq h(\mathbf{Z}|\theta)$$

$$= \Pr[\theta = 1] \cdot h(\mathbf{Z}|\theta = 1) + \Pr[\theta = 2] \cdot h(\mathbf{Z}|\theta = 2)$$
$$= \lambda \cdot \frac{1}{2} \log\Big((2\pi e)^n \det \mathsf{K}_1 \Big) + \bar{\lambda} \cdot \frac{1}{2} \log\Big((2\pi e)^n \det \mathsf{K}_2 \Big).$$

Combining these two inequalities and simplifying leads to

$$\log \det(\lambda \mathsf{K}_1 + \bar{\lambda} \mathsf{K}_2) \geq \log\Big[(\det \mathsf{K}_1)^{\lambda} \, (\det \mathsf{K}_2)^{\bar{\lambda}} \Big],$$

and exponentiating both sides of the inequality leads to the desired result.

Problem 2 — *Continuous Fano Inequality*

a) Let μ be the mean of X. We have

$$\begin{aligned}
\tau^2 &= \mathsf{E}\Big[(X - \hat{X})^2 \Big] \\
&= \mathsf{E}\Big[\big\{ (X - \mu) - (\hat{X} - \mu) \big\}^2 \Big] \\
&= \mathsf{E}\Big[(X - \mu)^2 \Big] - 2\, \mathsf{E}\Big[(X - \mu)(\hat{X} - \mu) \Big] + \mathsf{E}\Big[(\hat{X} - \mu)^2 \Big] \\
&\stackrel{\text{(i)}}{=} \mathsf{Var}(X) - 2 \underbrace{\mathsf{E}\Big[X - \mu \Big]}_{=0} \mathsf{E}\Big[\hat{X} - \mu \Big] + \underbrace{\mathsf{E}\Big[(\hat{X} - \mu)^2 \Big]}_{\geq 0} \\
&\geq \mathsf{Var}(X),
\end{aligned} \tag{1}$$

where (i) holds because X and $\hat{X}$ are independent. Since Gaussian random variables maximize differential entropy among all random variables having the same variance, we also have

$$h(X) \leq \frac{1}{2} \log_2 \Big(2\pi e \, \mathsf{Var}(X) \Big). \tag{2}$$

Combining (1) and (2) leads to the desired result:

$$\tau^2 \geq \frac{1}{2\pi e} 2^{2h(X)}.$$

Equality holds if and only if (1) and (2) both hold with equality, i.e., if and only if $\hat{X}$ is (deterministically) equal to μ and X is a Gaussian random variable.

b) We have

$$\begin{aligned}
\tau^2 &= \mathsf{E}\Big[(X - \hat{X})^2 \Big] \\
&\stackrel{\text{(i)}}{=} \int_{-\infty}^{\infty} f_Y(y) \cdot \mathsf{E}\Big[(X - \hat{X})^2 \Big| Y = y \Big] \, \mathrm{d}y \\
&\stackrel{\text{(ii)}}{\geq} \frac{1}{2\pi e} \int_{-\infty}^{\infty} f_Y(y) \cdot 2^{2h(X|Y=y)} \, \mathrm{d}y \\
&\stackrel{\text{(iii)}}{\geq} \frac{1}{2\pi e} 2^{\int_{-\infty}^{\infty} f_Y(y) \cdot 2h(X|Y=y) \, \mathrm{d}y} \\
&\stackrel{\text{(iv)}}{=} \frac{1}{2\pi e} 2^{2h(X|Y)},
\end{aligned}$$

where (i) follows from the law of total expectation; (ii) follows from Part a) because conditional on $Y = y$, X and $\hat{X}$ are independent; (iii) follows from Jensen's inequality because the function $u \mapsto 2^u$ is convex; and (iv) follows from the definition of $h(X|Y)$.

Problem 3 ***Maximum Differential Entropy***

According to the maximum differential entropy principle, if a density of the form

$$f(x) = e^{\lambda_0+\lambda_1 x+\lambda_2 \ln x}\, \mathrm{I}\{x > 0\}$$

satisfying the constraints exists, it uniquely maximizes differential entropy. Substituting

$$\begin{aligned}\lambda_0 &= \ln c,\\ \lambda_1 &= -\frac{1}{\theta},\\ \lambda_2 &= k-1,\end{aligned}$$

we obtain

$$f(x) = c \cdot e^{-\frac{x}{\theta}} x^{k-1}\, \mathrm{I}\{x > 0\},$$

which corresponds to a *gamma distribution* with shape k and scale θ. Looking up the properties of the gamma distribution, we obtain

$$\begin{aligned}c &= \frac{1}{\theta^k \Gamma(k)},\\ \mathsf{E}[X] &= k \cdot \theta \overset{!}{=} \alpha_1, &\quad (3)\\ \mathsf{E}[\ln X] &= \psi(k) + \ln\theta \overset{!}{=} \alpha_2, &\quad (4)\end{aligned}$$

where $\Gamma(\cdot)$ is the gamma function and $\psi(\cdot)$ is the digamma function. From (3) and (4) it is possible to determine the parameters k and θ, and the maximum differential entropy density then is

$$f(x) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}\, \mathrm{I}\{x > 0\}.$$

Problem 4 ***Minimum $D(P\|Q)$ under Constraints on P***

a) We form the functional

$$\mathrm{J}(P) = \sum_x P(x) \ln \frac{P(x)}{Q(x)} + \mu_0 \left(1 - \sum_x P(x)\right) + \sum_{i=1}^{\infty} \mu_i \left(\alpha_i - \sum_x P(x) g_i(x)\right).$$

We then compute its partial derivative with respect to $P(\hat{x})$:

$$\frac{\partial \mathrm{J}(P)}{\partial P(\hat{x})} = \ln \frac{P(\hat{x})}{Q(\hat{x})} + 1 - \mu_0 - \sum_{i=1}^{\infty} \mu_i g_i(\hat{x}).$$

Setting this partial derivative to zero, we guess that the PMF minimizing $D(P||Q)$ subject to the constraints should have the form

$$\begin{aligned}P^*(x) &= Q(x)\, e^{-1+\mu_0+\sum_{i=1}^{\infty} \mu_i g_i(x)}\\ &= Q(x)\, e^{\lambda_0+\sum_{i=1}^{\infty} \lambda_i g_i(x)}, &\quad (5)\end{aligned}$$

where we used the substitutions $\lambda_0 = -1 + \mu_0$ and $\lambda_i = \mu_i, i = 1, 2, \ldots$.

b) We next verify that P^* given in (5) indeed minimizes $D(P\|Q)$ if it satisfies all the constraints. To this end, let P be a PMF satisfying all the constraints. Then,

$$\begin{aligned}
D(P\|Q) &= \sum_x P(x)\log\frac{P(x)\,P^*(x)}{Q(x)\,P^*(x)} \\
&= \sum_x P(x)\log\frac{P(x)}{P^*(x)} + \sum_x P(x)\log\frac{P^*(x)}{Q(x)} \\
&= D(P\|P^*) + \sum_x P(x)\log\frac{P^*(x)}{Q(x)} \\
&\overset{\text{(i)}}{\geq} \sum_x P(x)\log\frac{P^*(x)}{Q(x)} \\
&= \sum_x P(x)\left(\lambda_0 + \sum_{i=1}^{\infty}\lambda_i g_i(x)\right)\log e \\
&= \left(\lambda_0\sum_x P(x) + \sum_{i=1}^{\infty}\lambda_i\sum_x P(x)g_i(x)\right)\log e \\
&\overset{\text{(ii)}}{=} \left(\lambda_0\sum_x P^*(x) + \sum_{i=1}^{\infty}\lambda_i\sum_x P^*(x)g_i(x)\right)\log e \\
&= \sum_x P^*(x)\log\frac{P^*(x)}{Q(x)} \\
&= D(P^*\|Q),
\end{aligned}$$

where (i) holds because relative entropy is nonnegative; and (ii) holds because both P and P^* satisfy the constraints. We thus conclude that P^* indeed minimizes $D(P\|Q)$.

Problem 5 — *Maximum Conditional Differential Entropy*

We have

$$\begin{aligned}
h(X|Y) &\overset{\text{(i)}}{=} h(X-\alpha Y|Y) \\
&\overset{\text{(ii)}}{\leq} h(X-\alpha Y) \\
&\overset{\text{(iii)}}{\leq} \frac{1}{2}\log\Bigl(2\pi e\,\mathsf{Var}(X-\alpha Y)\Bigr) \\
&\overset{\text{(iv)}}{=} \frac{1}{2}\log\Bigl(2\pi e\,\mathsf{E}\bigl[(X-\alpha Y)^2\bigr]\Bigr) \\
&= \frac{1}{2}\log\Bigl(2\pi e\bigl(\mathsf{E}\bigl[X^2\bigr] - 2\alpha\,\mathsf{E}[XY] + \alpha^2\,\mathsf{E}\bigl[Y^2\bigr]\bigr)\Bigr) \\
&\overset{\text{(v)}}{=} \frac{1}{2}\log\Bigl(2\pi e\bigl(\sigma_X^2 - 2\alpha\rho\sigma_X\sigma_Y + \alpha^2\sigma_Y^2\bigr)\Bigr),
\end{aligned}$$

where (i) holds because subtracting a constant from a random variable does not change its differential entropy; (ii) holds because conditioning does not increase entropy; (iii) holds because Gaussians maximize differential entropy among all random variables with the same variance; and (iv) and (v) hold because X and Y have zero mean. The choice $\alpha = \frac{\rho\sigma_X}{\sigma_Y}$ leads to the desired result:

$$h(X|Y) \leq \frac{1}{2}\log\Bigl(2\pi e(1-\rho^2)\sigma_X^2\Bigr).$$

Note that $X - \alpha Y$ and Y are uncorrelated with this choice of α:

$$\mathsf{E}[(X - \alpha Y)Y] = \mathsf{E}[XY] - \alpha\,\mathsf{E}\big[Y^2\big] = \rho\sigma_X\sigma_Y - \frac{\rho\sigma_X}{\sigma_Y}\sigma_Y^2 = 0.$$

Now assume that X and Y are jointly Gaussian. In this case, uncorrelatedness implies independence, and therefore (ii) holds with equality. Since a linear function of jointly Gaussian random variables is Gaussian, $X - \alpha Y$ is Gaussian and (iii) also holds with equality. We conclude that if X and Y are jointly Gaussian, we have

$$h(X|Y) = \frac{1}{2}\log\Big(2\pi e(1-\rho^2)\sigma_X^2\Big).$$

