Tsinghua-Berkeley Shenzhen Institute
Information Theory and Statistical Learning
Fall 2020

Problem Set 4

Issued: Monday 16th November, 2020 **Due:** Monday 30th November, 2020

Notations: We use $\mathsf{x}, \mathsf{y}, \mathsf{w}$ and $\underline{\mathsf{x}}, \underline{\mathsf{y}}, \underline{\mathsf{w}}$ to denote random variables and random vectors.

4.1. Please review Chapter 12 in Cover's book, then you can get some ideas on how to find the K-L divergence in Sanov's Theorem. Let x_i be i.i.d. $\sim \mathcal{N}(0, \sigma^2)$:

(a) Find the behavior of $-\frac{1}{n}\log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathsf{x}_i^2 \geq \alpha^2\right)$. This can be done from the first principles (since the normal distribution is nice) or by using Sanov's theorem.

(b) What does the data look like if $\frac{1}{n}\sum_{i=1}^n \mathsf{x}_i^2 \geq \alpha^2$. That is, what is the distribution that minimizes the K-L divergence in the Sanov's theorem.

Solution: A simple conclusion in Chapter 12 tells that the maximum entropy distribution is of the form $f(x) \sim Ce^{-\beta x^2}$ (Gaussian) and the constraint is $\mathbb{E}_f[\mathsf{x}^2] = \alpha^2$. Therefore it is $\mathcal{N}(0, \alpha^2)$.

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n \mathsf{x}_i^2 \geq \alpha^2\right) &= \operatorname*{arg\,min}_{P\in\{P:\mathbb{E}_P[\mathsf{x}^2]\geq\alpha^2\}} D(P\|Q) \\
&= D(\mathcal{N}(0,\alpha^2)\|\mathcal{N}(0,\sigma^2)) \\
&= \frac{1}{2}\frac{\alpha^2}{\sigma^2} + \frac{1}{2}\log\frac{\sigma^2}{\alpha^2}
\end{aligned}$$

4.2. We hope to derive an asymptotic value of $\binom{n}{k}$.

(a) Firstly, let's prove the lemma about Stirling's approximation of factorials, which we have used before.

$$\left(\frac{n}{e}\right)^n \leq n! \leq n\left(\frac{n}{e}\right)^n$$

Please justify the following steps:

$$\ln(n!) = \sum_{i=2}^{n-1} \ln i + \ln n \leq \cdots$$

$$\ln(n!) = \sum_{i=1}^{n} \ln i \geq \cdots$$

(b) If $0 < p < 1$, and $k = \lfloor np \rfloor$, i.e., k is the largest integer less than or equal to np, then please find

$$\lim_{n\to\infty} \frac{1}{n}\log\binom{n}{k}$$

Could you explain it without Stirling's Approximation?

Now let p_i's be a probability distribution on m symbols. Guess what is

$$\lim_{n\to\infty} \frac{1}{n} \log \begin{pmatrix} & & n & & \\ \lfloor np_1 \rfloor & \lfloor np_2 \rfloor & \cdots & \lfloor np_{m-1} \rfloor & (n - \sum_{i=1}^{m-1} \lfloor np_i \rfloor) \end{pmatrix}$$

Solution:

(a)

$$\ln(n!) = \sum_{i=2}^{n-1} \ln i + \ln n \le \int_2^{n-1} \ln x \mathrm{d}x + \ln n$$

$$\ln(n!) = \sum_{i=1}^{n} \ln i \ge \int_0^{n} \ln x \mathrm{d}x$$

Then you can get the approximation.

(b) By applying the approximation, we can easily get $\lim_{n\to\infty} \frac{1}{n} \log \binom{n}{k} = H(p) \triangleq -p \log p - (1-p)\log(1-p)$. Let's think about the conclusion, if we have a sequence with n i.i.d. rv from Bern(p), when n goes to ∞ with high probability, the sequence will have k '0's. For coding the sequence, we need at least $\log \binom{n}{k}$ bits, consistent with the entropy of the sequence $nH(p)$. That's what the result says.
The result is useful in the Sanov's theorem for Bernoulli rv's.

$$\lim_{n\to\infty} \frac{1}{n} \log \begin{pmatrix} & & n & & \\ \lfloor np_1 \rfloor & \lfloor np_2 \rfloor & \cdots & \lfloor np_{m-1} \rfloor & (n - \sum_{i=1}^{m-1} \lfloor np_i \rfloor) \end{pmatrix} = H(p_1, \cdots, p_m)$$

4.3. Consider the set of distributions on $\Omega = \{0, 1, 2\}$ and note that they lie on the 2-simplex

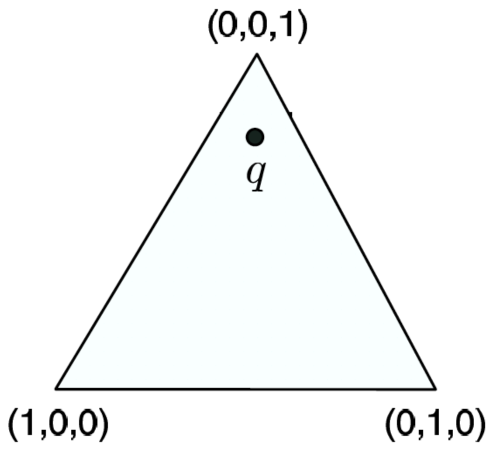
$$\{p = (p_0, p_1, p_2) : p_0 + p_1 + p_2 = 1, p_0 \ge 0, p_1 \ge 0, p_2 \ge 0\}$$

represented by the triangular figure. Let y be a random variable such that $p_{\mathsf{y}}(i) = p_i, i \in \{0, 1, 2\}$. Let $q = (1/6, 1/6, 2/3)$ be a particular probability mass function.

(a) Draw on the simplex the linear family corresponding to the expectation $\mathbb{E}[\mathsf{y}] = 0$, i.e. draw $\mathcal{L}_0 = \{p : \mathbb{E}_p[\mathsf{y}] = 0\}$.

(b) Draw $\mathcal{L}_{1/2} = \{p : \mathbb{E}_p[\mathsf{y}] = 1/2\}$

(c) Specify the exponential family $\mathcal{E}$ that passes through q and is orthogonal to $\mathcal{L}_{1/2}$, and draw the entire family on the 2-simplex.



Hint: Remember we introduced two versions of the exponential family, which are Lagrange-Multiplier induced one and parameterized one. You might be confused when you are facing cardinality-3 distributions, especially the Lagrange-Multiplier induced one. It is good if you can think about the equivalency of the two versions. Let's do the problem firstly under the parameterized version. That is $\mathcal{E} = \{\tilde{q} : \tilde{q} = qe^{sf(\mathsf{y})-\alpha(s)}\}$. Following the definition above, $f(\mathsf{y}) = \mathsf{y}$.

(d) Calculate the I-projection $p^\star$ of q onto $\mathcal{L}_{1/2}$ and mark it on the simplex.

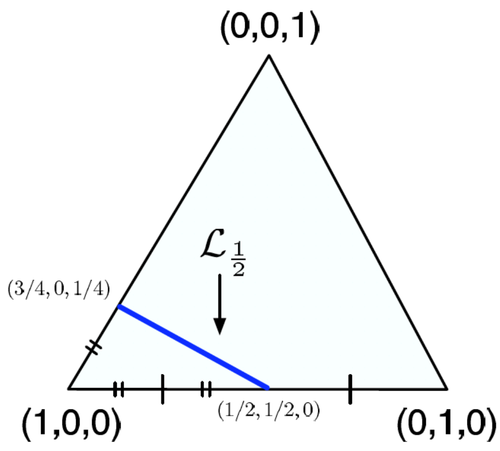
(e) Draw $\mathcal{P} = \{p : \mathbb{E}_p[\mathsf{y}] \leq 1/2\}$.

(f) Calculate the I-projection $p^\star$ of q onto $\mathcal{P}$ and mark it.
Hint: $D(\cdot\|q)$ is convex in its first argument.

Solution:

(a) Boundary point $(1, 0, 0)$. Figure ommited.

(b) We seek the set of points p such that $\mathbb{E}_p[\mathsf{y}] = 0p_0 + 1p_1 + 2p_2 = 1/2$. Letting $\lambda = 2p_1$, it follows that $\{(3-\lambda)/4, \lambda/2, (1-\lambda)/4\}$.

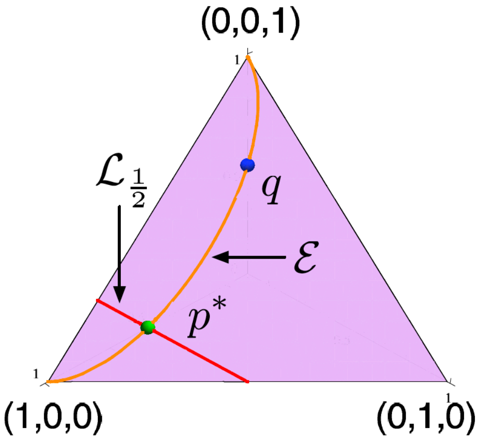


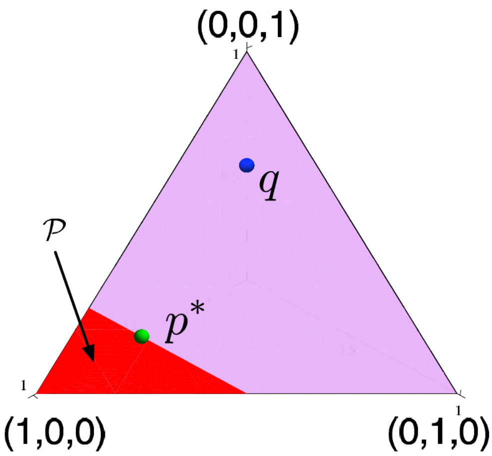
(c) The family is

$$\left(\frac{1}{1+e^x+4e^{2x}}, \frac{e^x}{1+e^x+4e^{2x}}, \frac{4e^{2x}}{1+e^x+4e^{2x}}\right).$$

Therefore, $p(0)p(2) = 4p^2(1) = 4(1p(0)p(2))^2$; the family is an ellipse.

(d) $p^\star = (2/3, 1/6, 1/6)$.





(e)

(f) The set $\mathcal{P}$ can be thought of as the collection of lines parallel to $\mathcal{L}_{1/2}$ of the form $L_\mu = p : \mathbb{E}_p[\mathsf{y}] = \mu$ for $\mu \in [0, 1/2]$. Let

$$p_\mu^\star \triangleq \arg\min_{p\in\mathcal{L}_\mu} D(p\|q)$$

$$p_\lambda = \lambda p_\mu^\star + (1-\lambda)q$$

where $\lambda \in [0,1]$.
Then $\mathbb{E}_{p=0}[\mathsf{y}] = \mathbb{E}_q[\mathsf{y}] = 3/2$, and $\mathbb{E}_{p=1}[\mathsf{y}] = E_{p_\mu^\star}[\mathsf{y}] = \mu \le 1/2$. By continuity we see that there exists a $\lambda^\star$ such that $\mathbb{E}_{p_{\lambda^\star}}[\mathsf{y}] = 1/2$ and hence $p_{\lambda^\star} \in \mathcal{L}_{1/2}$. Therefore,

$$D(p^\star\|q) \le D(p_{\lambda^\star}\|q) \le \lambda^] \star D(p_\mu^\star\|q) + (1-\lambda^\star)D(q\|q).$$

Hence for any given μ,

$$D(p^\star_\mu \| q) \geq \frac{1}{\lambda^\star} D(p^\star \| q) \geq D(p^\star \| q).$$

Hence, $p^\star = (2/3, 1/6, 1/6)$.

4.4. Let $q(y) > 0$ ($y = 0, 1, \cdots$) be a probability mass function for a random variable y and let $\mathcal{P}$ be the set of all PMFs defined over $\{0, \cdots, M-1\}$ for a known constant M:

$$\mathcal{P} \triangleq \{p : p(y) = 0,\ \forall y \geq M\}.$$

We can represent each element p of $\mathcal{P}$ as a M-dimensional vector $[p_0, \cdots, p_{M-1}]^{\mathrm{T}}$ that lies on a $(M-1)$-dimensional simplex, i.e., $\sum_{m=0}^{M-1} p_m = 1$.

(a) Show that, for all $p \in \mathcal{P}$, $D(q \| p) = \infty$

(b) Show that, for all $p \in \mathcal{P}$, $D(p \| q) < \infty$

(c) Find the I-projection of q onto $\mathcal{P}$, $p^\star = \arg\min_{p \in \mathcal{P}} D(p \| q)$, and the corresponding divergence $D(p^\star \| q)$ in terms of $Q(y) \triangleq \mathbb{P}(\mathsf{y} \leq y)$, the CDF of the random variable y.

Let $\mathcal{P}_\epsilon$ be the space of all PMFs with weight of ϵ on values M and above:

$$\mathcal{P}_\epsilon \triangleq \left\{ p : \sum_{y=M}^{\infty} p(y) = \epsilon \right\}$$

We can think of $\mathcal{P}_\epsilon$ as an extension of $\mathcal{P}$ to the distributions defined for all integers that only allows limited weight to be allocated to the values outside $\{0, \cdots, M-1\}$.

(d) Find the I-projection of q onto $\mathcal{P}_\epsilon$, $p^\star_\epsilon = \arg\min_{p \in \mathcal{P}_\epsilon} D(p \| q)$ and the corresponding divergence $D(p^\star_\epsilon \| q)$ in terms of $Q(y)$. Show that $\lim\limits_{\epsilon \to 0} D(p^\star_\epsilon \| q) = D(p^\star \| q)$.

(e) Show that $\mathcal{P}_\epsilon$ can be represented as a linear family of PMFs.

(f) Show that $p^\star_\epsilon$ belongs to an exponential family through q and find the value of the parameter that corresponds to $p^\star_\epsilon$.

Solution:

(a) ■

(b) ■

(c) We form the Lagrangian,

$$L = D(p \| q) - \lambda \left(\sum_{y=0}^{M-1} p(y) - 1 \right).$$

Then,
$$p^\star = q\exp(\lambda - 1).$$
According to the constaint, $\exp(\lambda-1) = \frac{1}{Q(M-1)}$ and $D(p^\star\|q) = \log\frac{1}{Q(M-1)}$.

(d) Follow the similar procedures,
$$p^\star_\epsilon = \begin{cases} q\dfrac{1-\epsilon}{Q(M-1)}, & y \le M-1 \\ q\dfrac{\epsilon}{1-Q(M-1)}, & y \ge M \end{cases}.$$
Then,
$$D(p^\star_\epsilon\|q) = (1-\epsilon)\log\frac{1-\epsilon}{Q(M-1)} + \epsilon\log\frac{\epsilon}{1-Q(M-1)},$$
and the limitation is easy to verify.

(e) Let $f(y) = \mathbb{1}_{y\ge M}$, then $\mathcal{P}_\epsilon = \{p : \mathbb{E}_p[f(\mathsf{y})] = \epsilon\}$.

(f) $\alpha(s) = \log\left(\sum_{y=0}^{M-1} q(y) + \sum_{y=M}^{\infty} q(y)e^s\right)$
$$p^\star_\epsilon = q\exp\left(\log\left(\frac{\epsilon}{1-\epsilon}\frac{Q(M-1)}{1-Q(M-1)}\right)\mathbb{1}_{y\ge M} - \log\frac{Q(M-1)}{1-\epsilon}\right)$$

4.5. *Joint Gaussian Distribution.* Suppose $\underline{\mathsf{x}} = (\mathsf{x}_1, \mathsf{x}_2)^{\mathrm{T}}$ is a Gaussian random vector with $\mathbb{E}[\mathsf{x}_1] = \mathbb{E}[\mathsf{x}_2] = 0, \mathrm{var}(\mathsf{x}_1) = \mathrm{var}(\mathsf{x}_2) = \sigma^2$, and $\rho_{\mathsf{x}} \triangleq \rho(\mathsf{x}_1, \mathsf{x}_2)$ denoting the correlation coefficient between x_1 and x_2. Let $\underline{\mathsf{y}} = (\mathsf{y}_1, \mathsf{y}_2)^{\mathrm{T}} \triangleq \mathbf{A}\underline{\mathsf{x}}$, where
$$\mathbf{A} = \begin{bmatrix} 1 & -\rho_{\mathsf{x}} \\ 0 & 1 \end{bmatrix}.$$
Then, $\underline{\mathsf{y}}$ is also a Gaussian random vector, since it is a linear transformation of $\underline{\mathsf{x}}$.

(a) Calculate $\mathbf{K}_{\mathsf{x}} \triangleq \mathrm{cov}(\underline{\mathsf{x}})$ and $\mathbf{K}_{\mathsf{y}} \triangleq \mathrm{cov}(\underline{\mathsf{y}})$.

(b) Prove that $\rho(\mathsf{y}_1, g(\mathsf{y}_2)) = 0$, for all functions[1] $g(\cdot)$. *Hint:* First prove that $\mathsf{y}_1 \perp\!\!\!\perp \mathsf{y}_2$.

(c) Prove that $\mathbb{E}[(\mathsf{x}_1 - \rho_{\mathsf{x}}\mathsf{x}_2)^2] \le \mathbb{E}[(\mathsf{x}_1 - g(\mathsf{x}_2))^2]$, for all functions $g\colon \mathbb{R}\to\mathbb{R}$. *Hint:* Rewrite the inequality using y_1 and y_2.

Solution:

(a)
$$\mathbf{K}_{\mathsf{x}} = \mathrm{cov}(\underline{\mathsf{x}}) = \mathbb{E}\left[\underline{\mathsf{x}}\underline{\mathsf{x}}^{\mathrm{T}}\right] - \mathbb{E}[\underline{\mathsf{x}}]\,\mathbb{E}[\underline{\mathsf{x}}]^{\mathrm{T}} = \mathbb{E}\left[\underline{\mathsf{x}}\underline{\mathsf{x}}^{\mathrm{T}}\right] = \sigma^2\begin{bmatrix} 1 & \rho_{\mathsf{x}} \\ \rho_{\mathsf{x}} & 1 \end{bmatrix}.$$

[1]Strictly speaking, $g(\cdot)$ is required to be measurable.

Then,

$$\mathbf{K}_{\mathsf{y}} = \mathbf{A}\mathbf{K}_{\mathsf{x}}\mathbf{A}^{\mathrm{T}} = \sigma^2 \begin{bmatrix} 1-\rho_{\mathsf{x}}^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

(b) Since $\mathbf{K}_{\mathsf{y}}$ is a diagonal matrix, we know that $\rho(\mathsf{y}_1, \mathsf{y}_2) = 0$. Then, as $\underline{\mathsf{y}}$ is a Gaussian vector, we obtain $\mathsf{y}_1 \perp\!\!\!\perp \mathsf{y}_2$, and thus the result follows.

(c) This is equivalent to $\mathbb{E}[\mathsf{y}_1^2] \leq \mathbb{E}\left[(\mathsf{y}_1 + \rho_{\mathsf{x}}\mathsf{y}_2 - g(\mathsf{y}_2))^2\right]$, which can be obtained from

$$\mathbb{E}\left[(\mathsf{y}_1 + \rho_{\mathsf{x}}\mathsf{y}_2 - g(\mathsf{y}_2))^2\right] = \mathbb{E}\left[\mathsf{y}_1^2\right] + \mathbb{E}\left[(\rho_{\mathsf{x}}\mathsf{y}_2 - g(\mathsf{y}_2))^2\right].$$

4.6. *Mathematical expectation and variance in estimation.* Suppose we want to estimate the value of y using an estimator $\hat{\mathsf{y}}$, and using its MSE (Mean Square Error) to evaluate the goodness of estimate, defined as

$$\mathrm{MSE}(\hat{\mathsf{y}}) \triangleq \mathbb{E}[(\mathsf{y} - \hat{\mathsf{y}})^2].$$

The estimator $\hat{\mathsf{y}}$ could be chosen from a set $\mathcal{A}$, and our goal is to find the best estimator in $\mathcal{A}$ which achieves the least MSE. Then the best estimator is called the MMSE (Minimum Mean Square Error) estimator.

(a) Assume we want to use a real number to estimate y, i.e., $\mathcal{A} = \mathbb{R}$.

i. Prove that $\mathbb{E}[\mathsf{y}]$ is the MMSE estimator:

$$\mathbb{E}[\mathsf{y}] = \operatorname*{arg\,min}_{\alpha \in \mathbb{R}} \mathbb{E}[(\mathsf{y} - \alpha)^2].$$

ii. Evaluate this estimator's MSE.

(b) Now you are allowed to use a function of x to estimate y, i.e., $\mathcal{A} = \{f(\cdot) : \mathcal{X} \mapsto \mathbb{R}\}$. Prove that:

i. $\mathbb{E}[\mathsf{y}|\mathsf{x}]$ is the MMSE estimator:

$$\mathbb{E}[\mathsf{y}|\mathsf{x}] = \operatorname*{arg\,min}_{f\colon \mathcal{X} \mapsto \mathbb{R}} \mathbb{E}[(\mathsf{y} - f(\mathsf{x}))^2],$$

ii. The MSE of estimator $\mathbb{E}[\mathsf{y}|\mathsf{x}]$ is

$$\mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) = \mathbb{E}[\mathrm{var}(\mathsf{y}|\mathsf{x})].$$

(c) Compare these two estimators. First, prove that

$$\mathsf{x} \perp\!\!\!\perp \mathsf{y} \implies \mathrm{MSE}(\mathbb{E}[\mathsf{y}]) = \mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) \implies \forall f,\ \rho(f(\mathsf{x}), \mathsf{y}) = 0,$$

where $\rho(\cdot,\cdot)$ is the Pearson correlation coefficient. In general, which one of these two estimators would have less MSE than the other?

Solution: (a)(b) can be solved by direct computations. For (c), note that we have (cf. PS2.1 (a) iv)

$$\begin{aligned}\mathrm{MSE}(\mathbb{E}[\mathsf{y}]) = \mathrm{var}(\mathsf{y}) = \mathbb{E}[\mathrm{var}(\mathsf{y}|\mathsf{x})] + \mathrm{var}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) &= \mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) + \mathrm{var}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) \\ &\geq \mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]).\end{aligned}$$

Therefore,

$$\mathrm{MSE}(\mathbb{E}[\mathsf{y}]) = \mathrm{MSE}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) \implies \mathrm{var}(\mathbb{E}[\mathsf{y}|\mathsf{x}]) = 0 \implies \mathbb{E}[\mathsf{y}|\mathsf{x}] = \mathbb{E}[\mathbb{E}[\mathsf{y}|\mathsf{x}]] = \mathbb{E}[\mathsf{y}].$$

As a result,

$$\mathbb{E}[f(\mathsf{x})\mathsf{y}] = \mathbb{E}[f(\mathsf{x})\,\mathbb{E}[\mathsf{y}|\mathsf{x}]] = \mathbb{E}[f(\mathsf{x})\,\mathbb{E}[\mathsf{y}]] = \mathbb{E}[f(\mathsf{x})]\,\mathbb{E}[\mathsf{y}],$$

which implies $\rho(f(\mathsf{x}), \mathsf{y}) = 0$.

4.7. Consider the estimation of one-hot encoded vectors, where the settings are similar to those of Problem 3.3. In particular, suppose y takes values from $\mathcal{Y} = \{1, 2, \cdots, k\}$, then its one hot encoding is a k-dimensional vector defined as $\underline{\mathsf{y}} \triangleq (\mathbb{1}_{\mathsf{y}=1}, \mathbb{1}_{\mathsf{y}=2}, \cdots, \mathbb{1}_{\mathsf{y}=k})^{\mathrm{T}}$, i.e., $\underline{\mathsf{y}}$ is the i-th vector of the standard basis if $\mathsf{y} = i$.

Now, we would use $\underline{\hat{\mathsf{y}}}$ to estimate $\underline{\mathsf{y}}$, and use its MSE to evaluate the goodness of estimate. The MSE is defined similarly as the scalar case, except that the scalar quadratic operator is replaced by the ℓ_2 norm squared:

$$\mathrm{MSE}(\underline{\hat{\mathsf{y}}}) \triangleq \mathbb{E}[\|\underline{\mathsf{y}} - \underline{\hat{\mathsf{y}}}\|_2^2].$$

Again, the estimator $\underline{\hat{\mathsf{y}}}$ could be chosen from a set $\mathcal{A}$.

(a) Suppose we want to use a vector to estimate $\underline{\mathsf{y}}$, i.e., $\mathcal{A} = \mathbb{R}^k$. Prove that $\underline{P}_{\mathsf{y}}(\cdot)$ is the MMSE estimator:

$$\underline{P}_{\mathsf{y}}(\cdot) = \operatorname*{arg\,min}_{\underline{\alpha} \in \mathbb{R}^k} \mathbb{E}[\|\underline{\mathsf{y}} - \underline{\alpha}\|_2^2],$$

where $\underline{P}_{\mathsf{y}}(\cdot) \triangleq [P_{\mathsf{y}}(1), P_{\mathsf{y}}(2), \cdots, P_{\mathsf{y}}(k)]^{\mathrm{T}}$.

(b) Now you are allowed to use a multivariant function of x to estimate $\underline{\mathsf{y}}$, i.e., $\mathcal{A} = \{\underline{f} : \mathcal{X} \mapsto \mathbb{R}^k\}$. Prove that the MMSE estimator is $\underline{P}_{\mathsf{y}|\mathsf{x}}(\cdot|\mathsf{x})$:

$$\underline{P}_{\mathsf{y}|\mathsf{x}}(\cdot|\mathsf{x}) = \operatorname*{arg\,min}_{\underline{f}\colon \mathcal{X} \mapsto \mathbb{R}^k} \mathbb{E}[\|\underline{\mathsf{y}} - \underline{f}(\mathsf{x})\|_2^2],$$

where $\underline{P}_{\mathsf{y}|\mathsf{x}}(\cdot|\mathsf{x}) \triangleq [P_{\mathsf{y}|\mathsf{x}}(1|\mathsf{x}), P_{\mathsf{y}|\mathsf{x}}(2|\mathsf{x}), \cdots, P_{\mathsf{y}|\mathsf{x}}(k|\mathsf{x})]^{\mathrm{T}}$.

Solution: The MMSE estimators are the expectation and conditional expectation:

$$\underline{P}_{\mathsf{y}}(\cdot) = \mathbb{E}\left[\underline{\mathsf{y}}\right]$$

and

$$\underline{P}_{\mathsf{y}|\mathsf{x}}(\cdot|\mathsf{x}) = \mathbb{E}\left[\underline{\mathsf{y}}|\mathsf{x}\right].$$

4.8. The data $\mathsf{x}[n] = ar^n + \mathsf{w}[n]$ for $n = 0, \cdots, N-1$ are observed. The random variables $\mathsf{w}[0], \cdots, \mathsf{w}[N-1]$ are i.i.d. Gaussian random variables with zero mean and variance σ^2. r is a non-zero constant. Find the Cramér-Rao bound for a. Does an efficient estimator exist? If so, what is it and what is its variance?

Solution: The Cramér-Rao bound for a is

$$\lambda_e(a) \geq \begin{cases} \sigma^2 \frac{1-r^2}{1-r^{2N}}, & |r| \neq 1 \\ \frac{\sigma^2}{N}, & |r| = 1 \end{cases}.$$

The efficient estimator exists, given by

$$a_{\text{eff}}(\mathsf{x}) = \frac{\underline{r}^{\mathrm{T}}\underline{x}}{\underline{r}^{\mathrm{T}}\underline{r}},$$

with $\underline{x} = [x[0], \ldots, x[N-1]]^{\mathrm{T}}$ and $\underline{r} = [r^0, \ldots, r^{N-1}]^{\mathrm{T}}$. Since it is efficient, the variance is given by the Cramér-Rao bound

$$\lambda_{\text{eff}}(a) = \begin{cases} \sigma^2 \frac{1-r^2}{1-r^{2N}}, & |r| \neq 1 \\ \frac{\sigma^2}{N}, & |r| = 1 \end{cases}.$$