Solutions To Problems of Chapter 2

2.1. Derive the mean and variance for the binomial distribution.

Solution: For the mean value we have that

$$\begin{aligned}
\mathbb{E}[\mathrm{x}] &= \sum_{k=0}^{n} \frac{kn!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{kn!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{((n-1)-(k-1))!(k-1)!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{l=0}^{n-1} \frac{(n-1)!}{((n-1)-l)!l!} p^l (1-p)^{(n-1)-l} \\
&= np(p+1-p)^{n-1} = np. \qquad (1)
\end{aligned}$$

For the variance we have

$$\begin{aligned}
\sigma_x^2 &= \sum_{k=0}^{n} (k-np)^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} + \\
& \quad \sum_{k=0}^{n} (np)^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} - \\
& \quad 2np \sum_{k=0}^{n} k \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}, \qquad (2)
\end{aligned}$$

(3)

or

$$\sigma_x^2 = \sum_{k=0}^{n} k^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} + (np)^2 - 2(np)^2, \qquad (4)$$

However,

$$\sum_{k=0}^{n} k^2 \frac{(n-1)!}{(n-k)!k!} p^k (1-p)^{n-k} =$$
$$np \sum_{k=1}^{n} k \frac{n!}{\big((n-1)-(k-1)\big)!(k-1)!} p^{k-1}(1-p)^{(n-1)-(k-1)} =$$
$$np \sum_{l=0}^{n-1} (l+1) \frac{(n-1)!}{\big((n-1)-l\big)!l!} p^l (1-p)^{(n-1)-l} =$$
$$np + np(n-1)p, \tag{5}$$

which finally proves the result.

2.2. Derive the mean and the variance for the uniform distribution.

Solution: For the mean we have

$$\begin{aligned} \mu = \mathbb{E}[\mathrm{x}] &= \int_a^b \frac{1}{b-a} x dx \\ &= \frac{1}{b-a} \frac{b}{2}\Big|_a^b = \frac{b+a}{2}. \end{aligned} \tag{6}$$

For the variance, we have

$$\begin{aligned} \sigma_x^2 &= \frac{1}{b-a} \int_a^b (x-\mu)^2 dx = \frac{1}{b-a} \int_{a-\mu}^{b-\mu} y^2 dy \\ &= \frac{1}{b-a} \frac{y^3}{3}\Big|_{a-\mu}^{b-\mu} \\ &= \frac{1}{12}(b-a)^2. \end{aligned} \tag{7}$$

2.3. Derive the mean and covariance matrix of the multivariate Gaussian.

Solution: Without harming generality, we assume that $\boldsymbol{\mu} = \mathbf{0}$, in order to simplify the discussion. We have that

$$\frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \int_{-\infty}^{+\infty} \boldsymbol{x} \exp\Big(-\frac{1}{2}\boldsymbol{x}^T \Sigma^{-1} \boldsymbol{x}\Big) d\boldsymbol{x}, \tag{8}$$

which due to the symmetry of the exponential results in $\mathbb{E}[\mathbf{x}] = \mathbf{0}$.

For the covariance we have that

$$\int_{-\infty}^{+\infty} \exp\Big(-\frac{1}{2}\boldsymbol{x}^T \Sigma^{-1} \boldsymbol{x}\Big) d\boldsymbol{x} = (2\pi)^{l/2}|\Sigma|^{1/2}. \tag{9}$$

Following similar arguments as for the univariate case given in the text, we are going yo take the the derivative on both sides with respect to matrix Σ. Recall from linear algebra the following formulas.

$$\frac{\partial \text{Trace}\{AX^{-1}B\}}{\partial X} = -(X^{-1}BAX^{-1})^T, \quad \frac{\partial |X^k|}{\partial X} = k|X^k|X^{-T}.$$

Hence, taking the derivatives of both sides in (9) with respect to Σ we obtain,

$$\frac{1}{2}\int_{-\infty}^{+\infty} \exp\Big(-\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\Big)\big(\Sigma^{-1}\boldsymbol{x}\boldsymbol{x}^T\Sigma^{-1}\big)^T d\boldsymbol{x} = \frac{1}{2}(2\pi)^{l/2}|\Sigma|^{1/2}\Sigma^{-T}, \tag{10}$$

which then readily gives the result.

2.4. Show that the mean and variance of the beta distribution with parameters a and b are given by

$$\mathbb{E}[\mathrm{x}] = \frac{a}{a+b},$$

and

$$\sigma_x^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$

Hint: Use the property $\Gamma(a+1) = a\Gamma(a)$.

Proof: We know that

$$\text{Beta}(x|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}.$$

Hence

$$\mathbb{E}[\mathrm{x}] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 x x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)},$$

which, using the property $\Gamma(a+1) = a\Gamma(a)$, results in

$$\mathbb{E}[\mathrm{x}] = \frac{a}{a+b}. \tag{11}$$

For the variance we have

$$\mathbb{E}[(\mathrm{x}-\mathbb{E}[\mathrm{x}])^2] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \left(x - \frac{a}{a+b}\right)^2 x^{a-1}(1-x)^{b-1}dx, \tag{12}$$

or

$$\begin{aligned}\sigma_x^2 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 x^{a+1}(1-x)^{b-1}dx \\ &+ \frac{a^2}{(a+b)^2}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 x^{a-1}(1-x)^{b-1}dx \\ &- 2\frac{a}{a+b}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 x^{a}(1-x)^{b-1}dx,\end{aligned} \tag{13}$$

and following a similar path as the one adopted for the mean, it is a matter of simple algebra to show that

$$\sigma_x^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$

2.5. Show that the normalizing constant in the beta distribution with parameters a, b is given by

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}.$$

Proof: The beta distribution is given by

$$\text{Beta}(x|a,b) = Cx^{a-1}(1-x)^{b-1},\ 0 \leq x \leq 1. \tag{14}$$

Hence

$$C^{-1} = \int_0^1 x^{a-1}(1-x)^{b-1}dx. \tag{15}$$

Let

$$x = \sin^2\theta \Rightarrow dx = 2\sin\theta\cos\theta d\theta. \tag{16}$$

Hence

$$C^{-1} = 2\int_0^{\frac{\pi}{2}} (\sin\theta)^{2a-1}(\cos\theta)^{2b-1}d\theta. \tag{17}$$

Recall the definition of the gamma function

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx,$$

and set

$$x = y^2 \Rightarrow dx = 2ydy,$$

hence

$$\Gamma(a) = 2\int_0^\infty y^{2a-1}e^{-y^2}dy. \tag{18}$$

Thus

$$\Gamma(a)\Gamma(b) = 4\int_0^\infty \int_0^\infty x^{2a-1}y^{2b-1}e^{-(x^2+y^2)}dxdy. \tag{19}$$

Let

$$x = r\sin\theta, y = r\cos\theta \Rightarrow dxdy = rdrd\theta.$$

Hence

$$\Gamma(a)\Gamma(b) = 4\int_0^{\frac{\pi}{2}} \int_0^\infty r^{2(a+b)-1}e^{-r^2}(\sin\theta)^{2a-1}(\cos\theta)^{2a-1}drd\theta. \tag{20}$$

where integration over θ is in the interval $\left[0, \frac{\pi}{2}\right]$ to guarantee that x remains non-negative. From (20) we have

$$\Gamma(a)\Gamma(b) = \left(2\int_0^\infty r^{2(a+b)-1}e^{-r^2}dr\right)\left(2\int_0^{\frac{\pi}{2}}(\sin\theta)^{2a-1}(\cos\theta)^{2b-1}d\theta\right)$$
$$= \Gamma(a+b)C^{-1},$$

which proves the claim.

2.6. Show that the mean and variance of the gamma pdf

$$\text{Gamma}(x|a,b) = \frac{b^a}{\Gamma(a)}x^{a-1}e^{-bx},\ a, b, x > 0.$$

are given by

$$\mathbb{E}[\mathrm{x}] = \frac{a}{b},$$
$$\sigma_x^2 = \frac{a}{b^2}.$$

Proof: We have that

$$\mathbb{E}[\mathrm{x}] = \frac{b^a}{\Gamma(a)}\int_0^\infty x^a e^{-bx}dx.$$

Set $bx = y$. Then

$$\mathbb{E}[\mathrm{x}] = \frac{b^a}{\Gamma(a)}\frac{1}{b^{a+1}}\int_0^\infty y^a e^{-y}dy$$
$$= \frac{1}{b\Gamma(a)}\Gamma(a+1) = \frac{a\Gamma(a)}{b\Gamma(a)} = \frac{a}{b}.$$

For the variance, the following is valid

$$\sigma_x^2 = \mathbb{E}[(\mathrm{x} - \frac{a}{b})^2] = \frac{b^a}{\Gamma(a)}\Big\{\int_0^\infty x^{a+1}e^{-bx}dx$$
$$+ \frac{a^2}{b^2}\int_0^\infty x^{a-1}e^{-bx}dx - 2\int_0^\infty x^a e^{-bx}dx\Big\},$$

and following a similar path as before we obtain

$$\sigma_x^2 = \frac{a}{b^2}.$$

2.7. Show that the mean and variance of a Dirichlet pdf with K variables, $\mathrm{x}_k,\ k=1,2,\dots,K$ and parameters $a_k,\ k=1,2,\dots,K$, are given by

$$\mathbb{E}[\mathrm{x}_k] = \frac{a_k}{\overline{a}},\ k=1,2,\dots,K$$
$$\sigma_k^2 = \frac{a_k(\overline{a}-a_k)}{\overline{a}^2(1+\overline{a})},\ k=1,2,\dots,K,$$
$$\mathrm{cov}[\mathrm{x}_i\mathrm{x}_j] = -\frac{a_i a_j}{\overline{a}^2(1+\overline{a})},\ i\neq j,$$

where $\overline{a}=\sum_{k=1}^{K} a_k$.

Solution: Without harm of generality, we will derive the mean for x_K. The others are derived similarly. To this end, we have

$$p(x_1,x_2,\dots,x_{K-1}) = C\prod_{k=1}^{K-1} x_k^{a_k-1}\left(1-\sum_{k=1}^{K-1} x_k\right)^{a_K-1}$$

where

$$C=\frac{\Gamma(a_1+a_2+\ldots+a_K)}{\Gamma(a_1)\Gamma(a_2)\ldots\Gamma(a_K)}.$$

$$\mathbb{E}[\mathrm{x}_K] = C\int_0^1\cdots\int_0^1\left[\int_0^{1-\sum_{k=1}^{K-1}x_k} x_K p(x_1,\dots,x_{K-1},x_K)dx_K\right]dx_{K-1}\ldots dx_1$$
$$= C\int_0^1\cdots\int_0^1\left[\int_0^{1-\sum_{k=1}^{K-1}x_k} x_K\prod_{k=1}^{K-1}x_k^{a_k-1}\left(1-\sum_{k=1}^{K-1}x_k\right)^{a_K-1} dx_K\right]dx_{K-1}\ldots dx_1$$
$$= C\int_0^1\cdots\int_0^1\prod_{k=1}^{K-1}x_k^{a_k-1}\left(1-\sum_{k=1}^{K-1}x_k\right)^{a_K}\left[\int_0^{1-\sum_{k=1}^{K-1}x_k} dx_K\right]dx_{K-1}\ldots dx_1,$$

or

$$\mathbb{E}[\mathrm{x}_K] = C\int_0^1\cdots\int_0^1\prod_{k=1}^{K-1}x_k^{a_k-1}\left(1-\sum_{k=1}^{K-1}x_k\right)^{a_K} dx_{K-1}\ldots dx_1$$
$$= C\frac{\Gamma(a_1)\ldots\Gamma(a_K+1)}{\Gamma(a_1+a_2+\ldots+a_K+1)}$$
$$= C\frac{a_K\Gamma(a_1)\ldots\Gamma(a_K)}{(a_1+a_2+\ldots+a_K)\Gamma(a_1+a_2+\ldots+a_K)}$$
$$= \frac{a_K}{\overline{a}}.$$

In the sequel, we will show that

$$\mathbb{E}[\mathrm{x}_i\mathrm{x}_j] = -\frac{a_i a_j}{\overline{a}^2(\overline{a}+1)},\ i\neq j.$$

We derive it for the variables x_K and x_{K-1}, since any of the variables can be taken in place of x_K and x_{K-1}. Hence,

$$\begin{aligned}
\mathbb{E}[\mathrm{x}_{K-1}\mathrm{x}_K] &= C\int_0^1 \cdots \int_0^1 \left[\int_0^{1-\sum_{k=1}^{K-1} x_k} \left(\prod_{k=1}^{K-2} x_k^{a_k-1}\right) x_{K-1}^{a_{K-1}} x_K^{a_K}\, dx_K\right] dx_{K-1}\ldots dx_1 \\
&= C\int_0^1 \cdots \int_0^1 \prod_{k=1}^{K-2} x_k^{a_k-1} x_{K-1}^{a_{K-1}} \left[\int_0^{1-\sum_{k=1}^{K-1} x_k} x_K^{a_K}\, dx_K\right] dx_{K-1}\ldots dx_1 \\
&= \frac{C}{a_K+1}\int_0^1 \cdots \int_0^1 \prod_{k=1}^{K-2} x_k^{a_k-1} x_{K-1}^{a_{K-1}} \left(1-\sum_{k=1}^{K-1} x_k\right)^{a_K+1} dx_{K-1}\ldots dx_1 \\
&= \frac{C}{a_K+1}\frac{\Gamma(a_1)\ldots\Gamma(a_{K-2})\Gamma(a_{K-1}+1)\Gamma(a_K+2)}{\Gamma(a_1+\ldots+a_{K-2}+a_{K-1}+a_K+2)} \\
&= \frac{C}{a_K+1}\frac{a_K a_{K-1}\Gamma(a_1)\ldots\Gamma(a_K)(a_K+1)}{(1+a_1+\ldots+a_K)(a_1+\ldots+a_K)\Gamma(a_1+\ldots+a_K)}
\end{aligned}$$

or

$$\mathbb{E}[\mathrm{x}_{K-1}\mathrm{x}_K] = \frac{a_K a_{K-1}}{\overline{a}(1+\overline{a})}.$$

Thus in general,

$$\mathbb{E}[\mathrm{x}_i\mathrm{x}_j] = \frac{a_i a_j}{\overline{a}(1+\overline{a})}.$$

For the covariance, we have

$$\begin{aligned}
\mathrm{cov}[\mathrm{x}_i\mathrm{x}_j] &= \mathbb{E}\left[\mathrm{x}_i - \mathbb{E}[\mathrm{x}_i]\right]\mathbb{E}\left[\mathrm{x}_j - \mathbb{E}[\mathrm{x}_j]\right] \\
&= \mathbb{E}[\mathrm{x}_i\mathrm{x}_j] - \mathbb{E}[\mathrm{x}_i]\,\mathbb{E}[\mathrm{x}_j],
\end{aligned}$$

or

$$\begin{aligned}
\mathrm{cov}[\mathrm{x}_i\mathrm{x}_j] &= \frac{a_i a_j}{\overline{a}(1+\overline{a})} - \frac{a_i a_j}{\overline{a}^2} \\
&= \frac{a_i a_j \overline{a} - a_i a_j(1+\overline{a})}{\overline{a}^2(1+\overline{a})} = -\frac{a_i a_j}{\overline{a}^2(1+\overline{a})}.
\end{aligned}$$

2.8. Show that the sample mean, using N i.i.d drawn samples, is an unbiased estimator with variance that tends to zero asymptotically, as $N \longrightarrow \infty$.

Solution: From the definition of the sample mean we have

$$\mathbb{E}[\hat{\mu}_N] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\mathrm{x}_n] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\mathrm{x}] = \mathbb{E}[\mathrm{x}]. \tag{21}$$

For the variance we have,

$$\begin{aligned}
\sigma^2_{\hat{\mu}_N} &= \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathrm{x}_i - \mu\right)\left(\frac{1}{N}\sum_{j=1}^{N}\mathrm{x}_j - \mu\right)\right] \\
&= \mathbb{E}\left[\frac{1}{N^2}\left(\sum_{i=1}^{N}(\mathrm{x}_i - \mu)\sum_{j=1}^{N}(\mathrm{x}_j - \mu)\right)\right] \qquad (22) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[(\mathrm{x}_i - \mu)(\mathrm{x}_j - \mu)\right]. \qquad (23)
\end{aligned}$$

However, since the samples are i.i.d. drawn, the expected value of the product is equal to the product of the mean values, hence it is zero except for $i = j$, which then results in

$$\sigma^2_{\hat{\mu}_N} = \frac{1}{N}\sigma^2_x,$$

which proves the claim.

2.9. Show that for WSS processes

$$r(0) \geq |r(k)|, \forall k \in \mathbb{Z},$$

and that for jointly WSS processes,

$$r_u(0)r_v(0) \geq |r_{uv}(k)|, \ \forall k \in \mathbb{Z}.$$

Solution: Both properties are shown in a similar way. So, we are going to focus on the first one. Consider the obvious inequality,

$$\mathbb{E}[|\mathrm{u}_n + \lambda \mathrm{u}_{n-k}|^2] \geq 0,$$

or

$$\mathbb{E}[|\mathrm{u}_n|^2] + |\lambda|^2\,\mathbb{E}[|\mathrm{u}_{n-k}|^2] \geq \lambda^* r(k) + \lambda r^*(k),$$

or

$$r(0) + |\lambda|^2 r(0) \geq \lambda^* r(k) + \lambda r^*(k).$$

This is true for any λ, thus it will be true for $\lambda = \frac{r(k)}{r(0)}$. Substituting, we obtain

$$r(0) \geq \frac{|r(k)|^2}{r(0)},$$

which proves the claim.

Similar steps are adopted in order to prove the property for the cross-correlation.

2.10. Show that the autocorrelation of the output of a linear system, with impulse response, $w_n,\ n \in \mathbb{Z}$, is related to the autocorrelation of the input process, via,

$$r_d(k) = r_u(k) * w_k * w^*_{-k}.$$

Solution: We have that

$$\begin{aligned} r_d(k) &= \mathbb{E}[\mathrm{d}_n \mathrm{d}^*_{n-k}] = \mathbb{E}\left[\sum_i w^*_i \mathrm{u}_{n-i} \sum_j w_j \mathrm{u}^*_{n-k-j}\right] \\ &= \sum_i \sum_j w^*_i w_j\, \mathbb{E}[\mathrm{u}_{n-i}\mathrm{u}^*_{n-k-j}] \\ &= \sum_j w_j \sum_i w^*_i r(k+j-i). \end{aligned} \tag{24}$$

Set

$$h(n) := w_n * r_u(n). \tag{25}$$

Then we can write,

$$\begin{aligned} r_d(k) &= \sum_j w_j h(k+j) = \sum_j w_j h\big(-((-k)-j)\big) = w^*_{-k} * h(-(-k)) \\ &= w^*_{-k} * w_k * r_u(k), \end{aligned}$$

which proves the claim.

2.11. Show that

$$\ln x \le x - 1.$$

Solution: Define the function

$$f(x) = x - 1 - \ln x.$$

then

$$f'(x) = 1 - \frac{1}{x}, \quad \text{and} \quad f''(x) = \frac{1}{x^2}.$$

Thus $x = 1$ is a minimum, i.e.,

$$f(x) \ge f(1) = 1 - 1 - 0 = 0.$$

or

$$\ln x \le x - 1.$$

2.12. Show that

$$I(\mathrm{x};\mathrm{y}) \ge 0.$$

Hint: Use the inequality of Problem 2.11.

Solution: By the respective definition, we have that

$$\begin{aligned} -I(\mathrm{x};\mathrm{y}) &= -\sum_x \sum_y P(x,y) \log \frac{P(x|y)}{P(x)} \\ &= \log e \sum_x \sum_y P(x;y) \ln \frac{P(x)}{P(x|y)}, \end{aligned}$$

where we have used only terms where $P(x,y) \neq 0$. Taking into account the inequality, we have that

$$\begin{aligned} -I(\mathrm{x};\mathrm{y}) \leq \; & \log e \sum_x \sum_y P(x,y) \left\{ \frac{P(x)}{P(x|y)} - 1 \right\} = \\ & \log e \sum_x \sum_y \{ P(x)P(y) - P(x,y) \} . \end{aligned}$$

Note that the summation over the terms in the brackets is equal to zero, which proves the claim.

Note that if the random variables are independent, then $P(x) = P(x|y)$ and $I(\mathrm{x};\mathrm{y}) = 0$.

2.13. Show that if $a_i,\ b_i,\ i = 1, 2, \ldots, M$ are positive numbers, such as

$$\sum_{i=1}^{M} a_i = 1, \text{ and } \sum_{i=1}^{M} b_i \leq 1,$$

then

$$-\sum_{i=1}^{M} a_i \ln a_i \leq -\sum_{i=1}^{M} a_i \ln b_i.$$

Solution: Recalling the inequality from Problem 2.11, that

$$\ln \frac{b_i}{a_i} \leq \frac{b_i}{a_i} - 1,$$

or

$$\sum_{i=1}^{M} a_i \ln \frac{b_i}{a_i} \leq \sum_{i=1}^{M} (b_i - a_i) \leq 0,$$

which proves the claim and where the assumptions concerning a_i and b_i have been taken into account .

2.14. Show that the maximum value of the entropy of a random variable occurs if all possible outcomes are equiprobable.

Solution Let $p_i, \ i = 1, 2, \ldots, M$ be the corresponding probabilities of the M possible events. According to the inequality in Problem 2.13 form $b_i = 1/M$, we have,

$$-\sum_{i=1}^{M} p_i \ln p_i \leq \sum_{i=1}^{M} p_i \ln M,$$

or

$$-\sum_{i=1}^{M} p_i \ln p_i \leq \ln M.$$

Thus the maximum value of the entropy is $\ln M$, which is achieved if all probabilities are equal to $1/M$.

2.15. Show that from all the pdfs which describe a random variable in an interval $[a, b]$ the uniform one maximizes the entropy.

Solution: The Lagrangian of the constrained optimization task is

$$L(p(\cdot), \lambda) = -\int_{-\infty}^{+\infty} p(x) \ln p(x) dx + \lambda \Big(\int_{-\infty}^{+\infty} p(x) dx - 1 \Big).$$

According to the calculus of variations (for the unfamiliar reader, treat $p(x)$ as a variable and take derivatives under the integrals as usual) we take the derivative and set it equal to zero, resulting in

$$\ln p(x) = \lambda - 1.$$

Plugging it in the constrain equation, and performing the integration results in

$$p(x) = \frac{1}{b - a},$$

which proves the claim.