

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,

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

where the formula for the binomial expansion has been employed. For the variance we have,

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

or

$$\sigma_x^2 = \sum_{k=0}^n k^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} + (np)^2 - 2(np)^2, \tag{3}$$

However,

$$\begin{aligned}
 \sum_{k=0}^n k^2 \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} &= \\
 np \sum_{k=1}^n k \frac{(n-1)!}{((n-1)-(k-1))!(k-1)!} p^{k-1} (1-p)^{(n-1)-(k-1)} &= \\
 np \sum_{l=0}^{n-1} (l+1) \frac{(n-1)!}{((n-1)-l)!l!} p^l (1-p)^{(n-1)-l} &= \\
 np + np(n-1)p, &
 \end{aligned} \tag{4}$$

which finally proves the result.

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

Solution: For the mean we have

$$\begin{aligned}
 \mu = \mathbb{E}[x] &= \int_a^b \frac{1}{b-a} x dx \\
 &= \frac{1}{b-a} \frac{b^2}{2} \Big|_a^b = \frac{b+a}{2}.
 \end{aligned} \tag{5}$$

For the variance, we have

$$\begin{aligned}
 \sigma_x^2 &= \frac{1}{b-a} \int_a^b (x-\mu)^2 dx = \frac{1}{b-a} \int_{a-\mu}^{b-\mu} y^2 dy \\
 &= \frac{1}{b-a} \frac{y^3}{3} \Big|_{a-\mu}^{b-\mu} \\
 &= \frac{1}{12} (b-a)^2.
 \end{aligned} \tag{6}$$

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

$$\frac{1}{(2\pi)^{l/2} |\Sigma|^{1/2}} \int_{-\infty}^{+\infty} \mathbf{x} \exp \left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} \right) d\mathbf{x}, \tag{7}$$

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

For the covariance we have that

$$\int_{-\infty}^{+\infty} \exp \left(-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} \right) d\mathbf{x} = (2\pi)^{l/2} |\Sigma|^{1/2}. \tag{8}$$

Following similar arguments as for the univariate case given in the text, we are going to take 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 (8) with respect to Σ we obtain,

$$\frac{1}{2} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}\right) (\Sigma^{-1} \mathbf{x} \mathbf{x}^T \Sigma^{-1})^T d\mathbf{x} = \frac{1}{2} (2\pi)^{l/2} |\Sigma|^{1/2} \Sigma^{-T}, \quad (9)$$

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

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

and

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

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

Solution: We know that

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

Hence

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

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

$$\mathbb{E}[x] = \frac{a}{a+b}. \quad (10)$$

For the variance we have

$$\mathbb{E}[(x - \mathbb{E}[x])^2] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \left(x - \frac{a}{a+b}\right)^2 x^{a-1} (1-x)^{b-1} dx, \quad (11)$$

or

$$\begin{aligned} \sigma_x^2 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^{a+1} (1-x)^{b-1} dx \\ &\quad + \frac{a^2}{(a+b)^2} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^{a-1} (1-x)^{b-1} dx \\ &\quad - 2 \frac{a}{a+b} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^a (1-x)^{b-1} dx, \end{aligned} \quad (12)$$

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

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

Solution: The beta distribution is given by

$$\text{Beta}(x|a, b) = Cx^{a-1}(1-x)^{b-1}, \quad 0 \leq x \leq 1. \quad (13)$$

Hence

$$C^{-1} = \int_0^1 x^{a-1}(1-x)^{b-1} dx. \quad (14)$$

Let

$$x = \sin^2 \theta \Rightarrow dx = 2 \sin \theta \cos \theta d\theta. \quad (15)$$

Hence

$$C^{-1} = 2 \int_0^{\frac{\pi}{2}} (\sin \theta)^{2a-1} (\cos \theta)^{2b-1} d\theta. \quad (16)$$

Recall the definition of the gamma function

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

and set

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

hence

$$\Gamma(a) = 2 \int_0^\infty y^{2a-1} e^{-y^2} dy. \quad (17)$$

Thus

$$\Gamma(a)\Gamma(b) = 4 \int_0^\infty \int_0^\infty x^{2a-1} y^{2b-1} e^{-(x^2+y^2)} dx dy. \quad (18)$$

Let

$$x = r \sin \theta, y = r \cos \theta \Rightarrow dx dy = r dr d\theta.$$

Hence

$$\Gamma(a)\Gamma(b) = 4 \int_0^{\frac{\pi}{2}} \int_0^\infty r^{2(a+b)-1} e^{-r^2} (\sin \theta)^{2a-1} (\cos \theta)^{2b-1} dr d\theta. \quad (19)$$

where integration over θ is in the interval $[0, \frac{\pi}{2}]$ to guarantee that x remains non-negative. From (19) we have

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

which proves the claim.

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

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

are given by

$$\begin{aligned}\mathbb{E}[x] &= \frac{a}{b}, \\ \sigma_x^2 &= \frac{a}{b^2}.\end{aligned}$$

Solution: We have that

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

Set $bx = y$. Then

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

For the variance, the following is valid

$$\begin{aligned}\sigma_x^2 &= \mathbb{E}\left[\left(x - \frac{a}{b}\right)^2\right] = \frac{b^a}{\Gamma(a)} \left\{ \int_0^\infty x^{a+1} e^{-bx} dx \right. \\ &\quad \left. + \frac{a^2}{b^2} \int_0^\infty x^{a-1} e^{-bx} dx - 2\frac{a}{b} \int_0^\infty x^a e^{-bx} dx \right\},\end{aligned}$$

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, \mathbf{x}_k , $k = 1, 2, \dots, K$ and parameters a_k , $k = 1, 2, \dots, K$, are given by

$$\begin{aligned}\mathbb{E}[x_k] &= \frac{a_k}{\bar{a}}, \quad k = 1, 2, \dots, K \\ \sigma_k^2 &= \frac{a_k(\bar{a} - a_k)}{\bar{a}^2(1 + \bar{a})}, \quad k = 1, 2, \dots, K, \\ \text{cov}[x_i, x_j] &= -\frac{a_i a_j}{\bar{a}^2(1 + \bar{a})}, \quad i \neq j,\end{aligned}$$

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

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

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

$$\begin{aligned} \mathbb{E}[x_K] &= C \int_0^1 \dots \int_0^1 \left[\int_0^{1-\sum_{k=1}^{K-1} x_k} x_K p(x_1, \dots, x_{K-1}, x_K) dx_K \right] dx_{K-1} \dots dx_1 \\ &= C \int_0^1 \dots \int_0^1 \left[\int_0^{1-\sum_{k=1}^{K-1} x_k} x_K \prod_{k=1}^{K-1} x_k^{a_k-1} \left(1 - \sum_{k=1}^{K-1} x_k\right)^{a_K-1} dx_K \right] dx_{K-1} \dots dx_1 \\ &= C \int_0^1 \dots \int_0^1 \prod_{k=1}^{K-1} x_k^{a_k-1} \left(1 - \sum_{k=1}^{K-1} x_k\right)^{a_K} \left[\int_0^{1-\sum_{k=1}^{K-1} x_k} dx_K \right] dx_{K-1} \dots dx_1, \end{aligned}$$

or

$$\begin{aligned} \mathbb{E}[x_K] &= C \int_0^1 \dots \int_0^1 \prod_{k=1}^{K-1} x_k^{a_k-1} \left(1 - \sum_{k=1}^{K-1} x_k\right)^{a_K} dx_{K-1} \dots dx_1 \\ &= C \frac{\Gamma(a_1) \dots \Gamma(a_K + 1)}{\Gamma(a_1 + a_2 + \dots + a_K + 1)} \\ &= C \frac{a_K \Gamma(a_1) \dots \Gamma(a_K)}{(a_1 + a_2 + \dots + a_K) \Gamma(a_1 + a_2 + \dots + a_K)} \\ &= \frac{a_K}{\bar{a}}. \end{aligned}$$

In the sequel, we will show that

$$\mathbb{E}[x_i x_j] = -\frac{a_i a_j}{\bar{a}^2 (\bar{a} + 1)}, \quad i \neq j.$$

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,

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

or

$$\mathbb{E}[x_{K-1}x_K] = \frac{a_K a_{K-1}}{\bar{a}(1+\bar{a})}.$$

Thus in general,

$$\mathbb{E}[x_i x_j] = \frac{a_i a_j}{\bar{a}(1+\bar{a})}.$$

For the covariance, we have

$$\begin{aligned}
 \text{cov}[x_i x_j] &= \mathbb{E}[x_i - \mathbb{E}[x_i]] \mathbb{E}[x_j - \mathbb{E}[x_j]] \\
 &= \mathbb{E}[x_i x_j] - \mathbb{E}[x_i] \mathbb{E}[x_j],
 \end{aligned}$$

or

$$\begin{aligned}
 \text{cov}[x_i x_j] &= \frac{a_i a_j}{\bar{a}(1+\bar{a})} - \frac{a_i a_j}{\bar{a}^2} \\
 &= \frac{a_i a_j \bar{a} - a_i a_j (1+\bar{a})}{\bar{a}^2(1+\bar{a})} = -\frac{a_i a_j}{\bar{a}^2(1+\bar{a})}.
 \end{aligned}$$

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 $N \rightarrow \infty$.

Solution: From the definition of the sample mean we have

$$\mathbb{E}[\hat{\mu}_N] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}[x_n] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}[x] = \mathbb{E}[x]. \quad (20)$$

For the variance we have,

$$\begin{aligned}\sigma_{\hat{\mu}_N}^2 &= \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N x_i - \mu \right) \left(\frac{1}{N} \sum_{j=1}^N x_j - \mu \right) \right] \\ &= \mathbb{E} \left[\frac{1}{N^2} \left(\sum_{i=1}^N (x_i - \mu) \sum_{j=1}^N (x_j - \mu) \right) \right] \quad (21)\end{aligned}$$

$$= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \mathbb{E} [(x_i - \mu)(x_j - \mu)]. \quad (22)$$

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

$$\sigma_{\hat{\mu}_N}^2 = \frac{1}{N} \sigma_x^2,$$

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,

$$r_u(0)r_v(0) \geq |r_{uv}(k)|^2, \forall k \in \mathbb{Z}.$$

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

$$\mathbb{E}[|u_n - \lambda u_{n-k}|^2] \geq 0,$$

or

$$\mathbb{E}[|u_n|^2] + |\lambda|^2 \mathbb{E}[|u_{n-k}|^2] \geq \lambda^* r(k) + \lambda r^*(k),$$

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

$$\begin{aligned} r_d(k) &= \mathbb{E}[d_n d_{n-k}^*] = \mathbb{E} \left[\sum_i w_i^* u_{n-i} \sum_j w_j u_{n-k-j}^* \right] \\ &= \sum_i \sum_j w_i^* w_j \mathbb{E}[u_{n-i} u_{n-k-j}^*] \\ &= \sum_j w_j \sum_i w_i^* r_u(k + j - i). \end{aligned} \quad (23)$$

Set

$$h(n) := w_n * r_u(n). \quad (24)$$

Then we can write,

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

which proves the claim.

2.11. Show that

$$\ln x \leq x - 1.$$

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.,

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

or

$$\ln x \leq x - 1.$$

2.12. Show that

$$I(x; y) \geq 0.$$

Hint: Use the inequality of Problem 2.11.

Solution: By the respective definition, we have that

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

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

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

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 $I(x; y) = 0$.

2.13. Show that if $a_i, b_i, i = 1, 2, \dots, M$ 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 , $i = 1, 2, \dots, M$, be the corresponding probabilities of the M possible events. According to the inequality in Problem 2.13 for $b_i = 1/M$, 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

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

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.