Coursework

Assigment 1

Q1. Given that the both balls in the box are red we can use bayesian theorem to find the probability of this event ocurring:

$$P(RR \mid RRR) = \frac{P(RRR \mid RR)\, P(RR)}{P(RRR)}$$

In order to calculate this probabilty every other probability should be calculate individu.

$P(RR) = \frac{1}{4}$ since there are only 4 possible outcomes.

$P(RRR \mid RR) = 1$

For the probability of having drawn 3 red balls we should calculate the probability of this event happen given the 3 different possibilities (Probality of having red and white RW, probability of having Red and red RR and probability of having white and red).

$$P(RRR) = P(RRR \mid RW) \cdot P(RW) \cdot 2 + P(RRR \mid RR) \cdot P(RR) =$$

$$\left(\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{4}\right) 2 + 1 \cdot \frac{1}{4}$$

$$\frac{1}{16} + \frac{1}{4} = \frac{5}{16}$$

The probability of drawning 3 red balls given that there are 2 red balls in the box

is given by

Now it is possible to calulate the probability of the event happen

$$P(RR \mid RRR) = \frac{P(RRR \mid RR)\, P(RR)}{P(RRR)} = \frac{1 \cdot \frac{1}{4}}{\frac{5}{16}} = \frac{4}{5}$$

Q2. The probability density function (pdf) of normal distribution is

$$f(x_i, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$-\infty < x < \infty$$

The likelihood function is:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2)$$

$$= (\sigma^2)^{-n/2}(2\pi)^{-n/2} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$\log L(\mu,\sigma^2) = \frac{-n}{2}\log\sigma^2 - \frac{n}{2}\log(2\pi) - \frac{\sum(x_i-\mu)^2}{2\sigma^2}$$

In order to find the maximum likelihood ~~it~~ is neccessary to derivate respect to μ and σ^2

with respect to μ

$$\frac{\partial \log L(\mu,\sigma^2)}{\partial\mu} = 0 + 0 - \frac{2\sum x_i - \mu(-1)}{2\sigma^2} = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - \mu = 0 \qquad \Rightarrow (x_1-\mu)+(x_2-\mu)+\dots+(x_n-\mu)=0$$

$$\Rightarrow (x_1+x_2+\dots+x_n) - n\mu = 0 \qquad = \sum_{i=1}^{n} x_i - n\mu = 0$$

$$n\mu = \sum_{i=1}^{n} x_i \qquad \mu = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \Rightarrow \mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

with respect to σ^2

$$\frac{\partial \log L(\mu,\sigma^2)}{\partial\sigma^2} = \frac{-n}{2\sigma^2} - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

$$= \frac{-n}{2\sigma^2} + \frac{\sum(x_i-\mu)^2}{2(\sigma^2)^2} = 0$$

$$= -n\sigma^2 + \sum(x_i-\mu)^2 = 0$$

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{n}$$

Knowing that $\mu = \frac{\sum_{i=1}^{n} x_i}{n}$ then $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu_{ML})^2}{N}$

Q2.

$$L = \frac{-N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2$$

In order to find the maximum likelihood the formula should be derivate and then equate it to 0.

$$\frac{dL}{d\mu} = 0 \qquad \frac{dL}{d\sigma^2} = 0$$

Q3. $E[X] = \int_{-\infty}^{\infty} N(x|\mu,\sigma^2) x \, dx = \mu$

$E[X^2] = \int_{-\infty}^{\infty} N(x|\mu,\sigma^2) x^2 dx = \mu^2 \, \sigma^2$

Since x_n and x_m are independent data points sampled from a Gaussian distribution and they both have mean μ and variance σ^2

$E[x_n x_m] = E[x_n] \cdot E[x_m] = \mu^2$

when $n=m$ the equation above applies

$E[x_n x_m] = E[x_n] \cdot E[x_n] = \mu^2 + I_{nm} \sigma^2$ so whenever $n \neq m$ $I_{nm} = 0$

$\mu_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_n$

$E(\mu_{ML}) = E\left[\frac{\sum_{i=1}^{N} x_n}{N}\right] = \frac{1}{N} E[\sum_{n=1}^{N} x_n] = \frac{1}{N} E[N \cdot x_n] = E[x_n] = \mu$

$\sigma^2_{ML} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{ML})^2$

$E[\sigma^2_{ML}] = \frac{1}{N} E[\sum_{n=1}^{N} (x_n - \mu_{ML})^2] = \frac{1}{N} E[\sum_{n=1}^{N} (x^2 - 2x_n \mu_{ML} + \mu^2_{ML})]$

$= \frac{1}{N} E[\sum_{n=1}^{N} x_n^2] - \frac{1}{N} E[\sum_{n=1}^{N} 2x_n \mu_{ML}] + \frac{1}{N} E[\sum_{n=1}^{N} \mu^2_{ML}]$

$\Rightarrow \frac{1}{N} E[\sum_{n=1}^{N} x_n^2] = \frac{1}{N} E[N \cdot x^2] = \mu^2 + \sigma^2$

$\Rightarrow -\frac{1}{N} E[\sum_{n=1}^{N} 2x_n \mu_{ML}] = \frac{-2}{N} E[\sum_{n=1}^{N} x_n (\frac{1}{N} \sum_{n=1}^{N} x_n)] = \frac{-2}{N^2} E[\sum_{n=1}^{N} x_n (\sum_{n=1}^{N} x_n)] = \frac{-2}{N^2} E[(\sum_{n=1}^{N} x_n)^2]$

$\Rightarrow E[\mu^2_{ML}] = E[(\frac{1}{N} \sum_{n=1}^{N} x_n)^2] = \frac{1}{N^2} E[(\sum_{n=1}^{N} x_n)^2]$

$E[\sigma^2_{ML}] = \mu^2 + \sigma^2 - \frac{2}{N^2} E[(\sum_{n=1}^{N} x_n)^2] + \frac{1}{N^2} E[(\sum_{n=1}^{N} x_n)^2] = \mu^2 + \sigma^2 - \frac{1}{N^2} E[(\sum_{n=1}^{N} x_n)^2]$

$\Rightarrow E[(\sum_{n=1}^{N} x_n)^2] = N^2 \mu^2 + N\sigma^2$

$E[\sigma^2_{ML}] = \mu^2 + \sigma^2 - \frac{1}{N^2}(N^2 \mu^2 + N\sigma^2) = \sigma^2 - \frac{1}{N}\sigma^2 = \left(\frac{N-1}{N}\right)\sigma^2$

Q4. $P(x) = \begin{cases} \lambda e^{(-\lambda x)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ $\quad \lambda > 0$

The likelihood function is the product of densities:

$$L(\lambda; x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} f_X(x_j; \lambda)$$
$$= \prod_{j=1}^{n} \lambda \exp(-\lambda x_j)$$
$$= \lambda^n \exp(-\lambda \sum_{j=1}^{n} x_j)$$
$$= \lambda^n \exp(-\lambda x_1 - \lambda x_2 - \lambda x_3 \ldots - \lambda x_n)$$
$$= \lambda^n \exp(-\lambda \sum_{j=1}^{n} x_j)$$

log-likehood

$$= \ln(L(\lambda; x_1, x_2, \ldots, x_n))$$
$$= \ln(\lambda^n) + \ln(\exp(-\lambda \sum_{j=1}^{n} x_j))$$
$$= n \ln(\lambda) + (-\lambda \sum_{j=1}^{n} x_j)$$
$$= n \ln(\lambda) - \lambda \sum_{j=1}^{n} x_j$$

To find the maximum of this the function must be derivated and equal to 0,

$$\frac{d}{d\lambda}\left(n \ln(\lambda) - \lambda \sum_{j=1}^{n} x_j\right) = 0$$

$$\frac{n}{\lambda} - \sum_{j=1}^{n} x_j = 0$$

$$\lambda = \frac{n}{\sum_{j=1}^{n} x_j} = \frac{1}{\frac{\sum_{j=1}^{n} x_j}{n}}$$

} this is the mean so the parameter is estimated as the reciprocal of the sample mean

Qs. $X \sim N(\mu, \Sigma)$ $X \in \mathbb{R}^2$ $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

Knowing $N(X|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right]$

The pdf equation is given by

$$P(X_1, X_2) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right] \text{ where } D = 2$$

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2)$$

$$(X-\mu) = \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix}$$

$$|\Sigma|^{1/2} = \sigma_1\sigma_2\sqrt{(1-\rho^2)}$$

$$|\Sigma|^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

$$(X-\mu)^T = [(X_1-\mu_1) \; (X_2-\mu_2)]$$

$$P(X_1, X_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}[(X_1-\mu_1) \; (X_2-\mu_2)] \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix}\right]$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} [(X_1-\mu_1)\sigma_2^2 - \rho\sigma_1\sigma_2 \; -(X_2-\mu_2)\rho\sigma_1\sigma_2 + (X_2-\mu_2)\sigma_1^2] \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix}\right]$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2\sigma_1^2\sigma_2^2(1-\rho^2)} (X_1-\mu_1)^2\sigma_2^2 - X_1\mu_1\rho\sigma_1\sigma_2 - (X_2-\mu_2)\rho\sigma_1\sigma_2(X_1-\mu_1) + (X_2-\mu_2)^2\sigma_1^2\right]$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(X_1-\mu_1)^2}{\sigma_1^2} + \frac{(X_2-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(X_1-\mu_1)}{\sigma_1}\frac{(X_2-\mu_2)}{\sigma_2}\right)\right]$$

Q6.

$P(X_1, X_2) = N(X \mid \mu, \Sigma)$ $\quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \sigma_1\sigma_2 \begin{bmatrix} \frac{\sigma_1}{\sigma_2} & \rho \\ \rho & \frac{\sigma_2}{\sigma_1} \end{bmatrix}$

$\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2}$

Knowing that

$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

precision matrix is given by $\Lambda = \Sigma^{-1}$ $\quad \Lambda = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}$

$N(x \mid \mu, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$

from the above formula, the exponential term in this case can be expressed:

$-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)$ and this can be expanded as:

$$-\frac{1}{2}(x_1-\mu_1)^T \Lambda_{aa}(x_1-\mu_1) - \frac{1}{2}(x_1-\mu_1)^T \Lambda_{ab}(x_2-\mu_2) - \frac{1}{2}(x_2-\mu_2)^T \Lambda_{ba}(x_1-\mu_2)$$
$$-\frac{1}{2}(x_2-\mu_2)^T \Lambda_{bb}(x_2-\mu_2)$$

The mean can be calculated from the term first order in x_2 which is equal to $\Sigma_{a|b}^{-1}\mu_{a|b}$ and the second order term in x_2 when x_1 is constant represents the inverse of covariance matrix $\Sigma_{2|1}^{-1}$

$$-\frac{1}{2}(x_2-\mu_{b_2})^T \Lambda_{ba}(x_1-\mu_{a_1}) - \frac{1}{2}(x_2-\mu_{b_2})^T \Lambda_{bb}(x_2-\mu_{b_2})$$

the second order term in this case is $-\frac{1}{2}(x_2)^T \Lambda_{bb}(x_2)$ $\quad \Sigma_{2|1} = \Lambda_{bb}^{-1}$

the first order term is given as $x_2^T[\Lambda \;\; - \Lambda_{ba}(x_1-\mu_1)]$ and the coefficient should be equal to $\Sigma_{2|1}^{-1}\mu_{2|1}$

$$\mu_{2|1} = \Sigma_{2|1}[\Lambda_{bb}\mu_2 - \Lambda_{ba}(x_1-\mu_1)] = \Lambda_{bb}^{-1}[\Lambda_{bb}\mu_2 - \Lambda_{ba}(x_1-\mu_1)] = \mu_2 - \Lambda_{bb}^{-1}\Lambda_{ba}(x_1-\mu_1)$$

$$\Lambda = \frac{1}{\sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

$$\Sigma_{2|1} = \Lambda_{bb}^{-1} = \frac{\sigma_1^2 \sigma_2^2 (1-\rho^2)}{\sigma_1^2} = \sigma_2^2 (1-\rho)^2$$

$$\mu_{2|1} = \mu_2 - \Lambda_{bb}^{-1} \Lambda_{ba} (x_1 - \mu_1) = \mu_2 - \sigma_2^2 (1-\rho)^2 \cdot \frac{-\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2(1-\rho)^2} \cdot (x_1 - \mu_1)$$

$$= \mu_2 + \frac{\sigma_2 \rho}{\sigma_1}(x_1 - \mu_1)$$

Then $$p(x_2|x_1) = \frac{1}{\sqrt{2\pi \Sigma_{2|1}}} \exp\left[-\frac{1}{2}(x_2 - \mu_{2|1})^T \Sigma_{2|1}^{-1} (x_2 - \mu_{2|1})\right]$$

$$= \frac{1}{\sqrt{(2\pi(1-\rho^2)}\,\sigma_2} \exp\left[-\frac{1}{2}\left(x_2 - \mu_2 - \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)\right)^T \frac{1}{\sigma_2^2(1-\rho^2)} \left(x_2 - \mu_2 - \frac{\sigma_2 \rho}{\sigma_1}(x_1 - \mu_1)\right)\right]$$

if $\sigma_1 = \sigma_2 = 1$

$$p(x_2|x_1) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2}(x_2 - \mu_2 - (x_1 - \mu_1))^T \frac{1}{(1-\rho)^2} (x_2 - \mu_2 - \rho(x_1 - \mu_1))\right]$$

Q7.

Knowing that $f(x) = \frac{1}{b-a}$

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a} \times 2[b^2 - a^2] = \frac{b+a}{2} = \mu$$

For the variance

$$V(x) = E[x^2] - [E(x)]^2 = \frac{1}{b-a}\int_a^b x^2 dx - \left(\frac{b+a}{2}\right)^2 = \frac{b^3+a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 =$$

$$= \frac{4(b^3-a^3) - 3(b-a)(b+a)}{12(b-a)}$$

$$= \frac{4b^3 - 4a^3 - 3(b-a)(a^2+2ab+b^2)}{12(b-a)}$$

$$= \frac{4b^3 - 4a^3 - 3a^2b - 6ab^2 - 3b^3 + 3a^3 + 6a^2b + 3ab^2}{12(b-a)}$$

$$= \frac{b^3 + 3a^2b - 3ab^2 - a^3}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}$$

Q7.b

Multivariate Gaussian

$$N(x|\mu, \Sigma) = \frac{1}{2\pi \frac{D}{2} |\Sigma|^{1/2}} \exp\left(\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

$$E[x] = \int N(x|\mu, \Sigma)\, x\, dx$$

$$E[x] = \frac{1}{(2\pi)\frac{D}{2} |\Sigma|\frac{1}{2}} \int \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) x\, dx \qquad z = x-\mu \quad x = z+\mu$$

$$\qquad \qquad \qquad dx = dz$$

$$E[x] = \frac{1}{2\pi \frac{D}{2} |\Sigma| \frac{1}{2}} \int \exp\left(-\frac{1}{2} z^T \Sigma^{-1} z\right)(z+\mu)\, dz$$

$$= \frac{1}{2\pi \frac{D}{2} |\Sigma| \frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} z^T \Sigma^{-1} z\right) z\, dz + \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} z^T \Sigma^{-1} z\right) \mu\, dz$$

$$= \frac{\mu}{2\pi \frac{D}{2} |\Sigma| \frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} z^T \Sigma^{-1} z\right) dz$$

the reason for the normalized multivariable gaussian,

$$\frac{1}{2\pi \frac{D}{2} |\Sigma| \frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} z^T \Sigma^{-1} z\right) dz = 1 \qquad E[x] = \mu$$

D^2 has second order moments generated by $E[x_i, x_j]$ second order moment in multivariable gaussian that can be driven as matrix $E[x x^T]$ can be formed as:

$$E[x x^T] = \frac{1}{2\pi \frac{D}{2} |\Sigma| \frac{1}{2}} \int \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) x x^T dx$$

recalling $z = x-\mu \quad x = z+\mu$

$$E[x x^T] = \frac{1}{2\pi \frac{D}{2} |\Sigma| \frac{1}{2}} \int \exp\left(\frac{-1}{2} z^T \Sigma^{-1} z\right)(z+\mu)(z+\mu)^T dz$$

$$= \frac{1}{2\pi \frac{D}{2} |\Sigma| \frac{1}{2}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} z^T \Sigma^{-1} z\right)\left(z z^T + z\mu^T + \mu z + \mu\mu^T\right) dz$$

Due to symmetry, the conditions μz^T and $\mu' z$ will dissapeared, where condition $\mu\mu^T$ is constant and the last term zz^T it is nercessary to deal with. By using the eigenvector expansion of the convariance matrix, it is possible to derive

$$z = \sum_{j=1}^{D} y_j u_j \quad . \quad y_j = u_j z$$

$$E(xx^T) = \mu\mu^T + \frac{1}{2\pi^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}} \int \exp\left(\frac{-1}{2} z^T \Sigma^{-1} z\right) z z^T dz$$

$$E[xx^T] = \mu\mu^T + \frac{1}{2\pi^{\frac{D}{2}}|\Sigma|^{\frac{1}{2}}} \sum_{i=1}^{D}\sum_{j=1}^{D} u_i u_j \exp\left(-\sum_{k=1}^{D} \frac{y_k^2}{2\lambda_k}\right) y_i y_j \, dy$$

$$E[xx^T] = \mu\mu^T + \sum_{i=1}^{D} u_i u_j \lambda_i$$

$$E[xx^T] = \mu\mu^T + \Sigma$$

The covariance can be found by substracting the mean from $E[xx^T]$

$$\mathrm{cov}[x] = E\left((x - E[x])(x - E[x])^T\right)$$

$$\mathrm{cov}[x] = \Sigma$$