Advanced Machine Learning course, IDC

Assignment 1

Yaniv Tal - 031431166

Question 1

Given:

$P(Fair) = 0.999$
$P(Head|Fair) = 0.5$

$P(Forged) = 0.001$
$P(Head|Forged) = 0.9$

Definition: n = The number of times coin falls on "heads" out of 10 coin tosses

According to Bayes rule:

$$P(Forged|n=9) = \frac{P(n=9|Forged) \cdot P(Forged)}{P(n=9)}$$

$$P(Fair|n=9) = \frac{P(n=9|Fair) \cdot P(Fair)}{P(n=9)}$$

From binomial distribution:

$P(n=9|Forged) = \binom{10}{9} \cdot 0.9^9 \cdot 0.1 = 0.38742$
$P(n=9|Fair) = \binom{10}{9} \cdot 0.5^9 \cdot 0.5 = 0.009765625$

$P(n=9) = P(n=9|Forged) \cdot P(Forged) + P(n=9|Fair) \cdot P(Fair) = 0.010143279864$

Putting it together:

a. Chance this is a Forged coin: $P(Forged|n=9) = \dfrac{0.38742 \cdot 0.001}{0.010143279864} = 0.038195$

b. Chance this is a Fair coin: $P(Fair|n=9) = \dfrac{0.009765625 \cdot 0.999}{0.010143279864} = 0.961805$

Question 2

Definition: n=Random variable which represents the number of girls born in a family

According to the above, $n \sim Geom(p=0.5)$

The The expected number of girls per family in the country is: $E(n) = \dfrac{1-p}{p} = 1$

The number of boys per family is deterministic (always 1) for families that are done with the recreation phase.

Given the above:

- **If we look only at families which already finished making kids - We should expect equal numbers of boys and girls**.
- **If we look at the entire population, including families which are in the middle of the recreational process** - we should expect slightly more girls, because there will be families with girls only who did not make a boy just yet.

Question 3

Using log likelihood:

$$\theta_{ML} = \underset{\theta\in\Omega}{argmax}\ LL(\theta) =$$
$$\underset{\theta\in\Omega}{argmax}\ log(P(x_1, \ldots, x_n|\theta)) =$$
$$\underset{\theta\in\Omega}{argmax}\ log \prod_{i=1}^{n} (P(x_i|\theta)) =$$
$$\underset{\theta\in\Omega}{argmax} \sum_{i=1}^{n} log(P(x_i|\theta)) =$$

a. Binomial MLE:

Suppose we got "Heads" k times in our sample.

$$LL(\theta) = logP(x_1 \ldots x_n|\theta) = log(p^k \cdot (1-p)^{n-k}) = k \cdot log(p) + (n-k) \cdot log(1-p)$$

Differentiate and set to Zero:

$$\frac{dL(\theta)}{dp} = \frac{k}{p} - \frac{n-k}{1-p} = 0$$
$$p = \frac{k}{p}$$

b. Poisson MLE:

$$L(\theta) = L(\lambda|x_1 \ldots x_n) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{k_i}}{k_i!}$$
$$LL(\theta) = lnL(\theta) = ln\left(\prod_{i=1}^{n} e^{-\lambda} \frac{\lambda k_i}{k_i!} \right) = \sum_{i=1}^{n} ln\left(e^{-\lambda} \frac{\lambda^{k_i}}{k_i!} \right) = -\lambda n - \sum_{i=1}^{n} ln(x_i!) + ln\lambda \sum_{i=1}^{n} x_i$$

Differentiate and set to Zero:

$$\frac{dL(\theta)}{d\lambda} = -m + \frac{1}{\lambda} \sum_{i=1}^{n} x_i = 0$$
$$\lambda = \frac{1}{n} \sum_{i=1}^{n} x_i$$

c. Normal MLE:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$
$$LL(\theta) = -n \cdot ln(\sigma\sqrt{2\pi}) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

Differentiate and set each parameter μ, σ to Zero:

$$\frac{\partial L(\theta)}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} = 0$$
$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\frac{\partial L(\theta)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^3} = 0$$
$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

Question 4

Section a.

To obtain the marginal distribution of each random variable, we need to integrate the bivariate term on the other variable in the range $[-\infty,\infty]$.
Since the bivariate Normal density function is symmetric for the two variables, I will integrate explicitly for one, the other will be the same with the opposite parameters.

The bivariate normal distribution density:

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

The marginal density $m(x_1)$ is the integral of f on x_2 from $-\infty$ to ∞:

$$m(x_1)=\int_{-\infty}^{\infty}f(x_1,x_2)dx_2=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}dx_2=$$

Taking constants elements out of the integral:

$$\frac{exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\}dx_2=$$

Complete to square (Algebra):

$$\frac{exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-\rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]\right\}dx_2=$$

Take the new independent element out of the integral:

$$\frac{exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+\frac{\rho^2}{2(1-\rho^2)}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]\right\}dx_2=$$

Simplify:

$$\frac{exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_2-\mu_2}{\sigma_2}\right)-\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right]^2\right\}dx_2$$

Define:

$$\upsilon:=\frac{1}{\sqrt{1-\rho^2}}\left[\frac{x_2-\mu_2}{\sigma 2}-\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right]$$

So we get:

$$d\upsilon=\frac{1}{\sqrt{1-\rho^2}\sigma_2}dx_2$$

Now we can introduce υ instead of x_2 into the equations of $m(x_1)$ and continue from the last step:

$$m(x_1)=\frac{exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}exp\left(-\frac{1}{2}\upsilon^2\right)\sqrt{1-\rho^2}\sigma_2 d\upsilon=$$

$$\frac{exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}{2\pi\sigma_1}\int_{-\infty}^{\infty}exp\left(-\frac{1}{2}\upsilon^2\right)d\upsilon$$

We can now use this integral:

$$\int_{-\infty}^{\infty}\frac{exp\left(-\frac{1}{2}\upsilon^2\right)}{\sqrt{2\pi}}=1\Longrightarrow\int_{-\infty}^{\infty}exp\left(-\frac{1}{2}\upsilon^2\right)=\sqrt{2\pi}$$

Introducing the integral into the last step:

$$m(x_1) = \frac{exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)}{2\pi\sigma_1}\sqrt{2\pi} =$$

$$\frac{1}{\sqrt{2\pi}\sigma_1}exp\left(-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right)$$

And as expected, we got the monovariate normal distribution density function.

Section b.

$$f_{x_1|x_2}(x_1) = \frac{f(x1,x2)}{f_{x_2}(x2)} =$$

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

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

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

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

Take $\frac{1}{\sigma_1^2}$ out of the brackets:

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

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

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

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

As required - This is the normal density function with parameters:

$$\mu = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2)$$
$$\sigma^2 = \sigma_1^2(1-\rho^2)$$

Question 5.

We can define an inner product on a set of random variables using the expected value of their product:
$\langle X, Y \rangle := E(XY)$

We can now use the above definition with the covariance definition:
$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \langle X - \mu_X, Y - \mu_y \rangle$

Cauchy-Schwartz inequality states that:
$|\langle u, v \rangle|^2 \leq \langle u, u \rangle \cdot \langle v, v \rangle$

Then, using Cauchy-Schwartz inequality with the covariance we get:
$|Cov(X, Y)|^2 = |\langle X - \mu_X, Y - \mu_Y \rangle|^2 \leq \langle X - \mu_X, X - \mu_x \rangle \langle Y - \mu_Y, Y - \mu_y \rangle = E((X - \mu_x)^2)E((Y - \mu_Y)^2) = Var(X)Var(Y)$

divide by the variances and take square root, we get:
$$\frac{|Cov(X, Y)|}{\sqrt{Var(X)}\sqrt{Var(Y)}} \leq 1$$

Note that the left side of the inequality is actually the definition of pearson ρ in absolute value. so we get:
$|\rho| \leq 1$

As required.