Exam about Information Field Theory

18.05.2015

Name: ____________________ Matrikelnummer: ____________________

- The exam consists of **nine exercises**. Please do check if you received all of them.
- In total one may gain 60 points. However it is likely to achieve the top grade with less points.
- The working time is **90 minutes**.
- **No aids** are allowed during the exam.

Question	**Points**
1	____/6
2	____/4
3	____/5
4	____/19
5	____/4
6	____/6
7	____/4
8	____/6
9	____/6
Bonus	
Total	____/60

Grade	

Recommendation: If you run into problems with one exercise, jump to the next one.
GOOD LUCK!

Important: In order to gain the full amount of possible points we strictly advise you to use conditional probabilities.

Question 1 ___/6

Prove or disprove the following relations:

a) $\overline{\overline{A} + \overline{B} + C} = A\overline{C} + B\overline{C}$ ___/2

b) $\overline{\overline{AB} + C} = \overline{A}C + \overline{B}C$ ___/2

c) $\left(\overline{A} + \overline{B}\right) C = \overline{AB} + \overline{C}$ ___/2

Question 2 ___/20

Let's play *White Wins*. This game is about drawing from an urn filled with a certain number of black and white balls. After each trial, when the ball's color has been identified, the ball is put pack into the urn.

Background information I: The pupil version of this game is played with one single urn u that is filled with an equal number of black and white balls; e.g., six each.

| ● ● ● ● ● ● ○ ○ ○ ○ ○ ○ |

a) Give the probability, $P(w_1|I)$, for drawing a white ball from the urn u. ___/1

b) Say you have drawn a white ball in the first trial, and put it back. Did you gain any knowledge in this first trail? What is the probability, $P(w_2|w_1, I)$, of drawing again a white ball from urn u under this condition? ___/2

c) Calculate the probability, $P(w_2, w_1|I)$, for drawing two white balls from the urn u sequentially, putting the first one back. ___/3

Background information J: The student version of *White Wins* is played with a set of $N + 1$ urns, $\{u_n\}_{0\leq n\leq N}$. Every urn u_n is filled with $N-n$ black and n white balls. $N = 3$ would then correspond to the same total numbers of balls as above.

| ● ● ● | | ● ● ○ | | ● ○ ○ | | ○ ○ ○ |

However, during the game one will only draw balls from a single urn that is randomly chosen prior to any trial; i.e., $P(u_n|J) = \frac{1}{N+1}$. All except the chosen urn are put aside.

d) Calculate the probability, $P(w_1|J)$, for drawing a white ball from the chosen urn. Compare your result with the one obtained in **a)**. ___/4

e) Say you have drawn a white ball in the first trial, and put it back. Did you gain any knowledge in this first trail? Infer the probability, $P(u_n|w_1, J)$, for the urn that might have been chosen, in general and in the case of $N = 3$. ___/4

Note: Your result of question **d)** might be useful for this.

f) Calculate the probability, $P(w_2, w_1|J)$, for drawing two white balls from the chosen urn sequentially, putting the first one back. Give a result for $N = 3$ and $N \to \infty$. Compare your result with the one obtained in **c)**. ___/6

Hint: Remember the finite sums,

$$\sum_{n=0}^{N} n = \frac{N(N+1)}{2} \qquad \sum_{n=0}^{N} n^2 = \frac{N(N+1)(2N+1)}{6} \qquad \sum_{n=0}^{N} n^3 = \left(\frac{N(N+1)}{2}\right)^3.$$

Question 3 ___/20

Consider a biased coin with the probability f of yielding head.

a) Calculate the probability of achieving exactly $k = 1$ heads after $n = 3$ tosses, given that $f = 1/3$. State the name of the probability distribution, which describes this stochastic process. ___/3

Hint:

$$\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!(n-k)!}$$

b) Now investigate the limit of a very large number of tosses, i.e. $n \to \infty$, while the expected number of heads $\lambda = nf$ remains finite. Under this condition, derive the probability distribution for yielding k-times head, given λ. How is the resulting probability distribution, given by ___/6

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!},$$

called?

Hint:

$$\frac{n!}{(n-k)!(n-\lambda)^k} \xrightarrow{n\to\infty} 1$$

c) Calculate the expectation value of k with the use of the probability distribution derived in **b)**. ___/4

d) Under the condition $\lambda \gg 1$, the probability distribution of **b)** makes a transition into a Gaussian distribution. Derive this Gaussian distribution $P(k|\lambda)$. For this purpose use the substitution ___/7

$$k = \lambda(1+\delta), \text{ with } \lambda \gg 1, \ \delta \ll 1,$$

and use Stirling's formula for $k!$:

$$k! \approx \sqrt{2\pi k}\left(\frac{k}{e}\right)^k .$$

Hint: Use $(1+\delta)^{\lambda(1+\delta)+1/2} \approx \exp\left(\lambda\delta + \lambda\delta^2/2 + \mathcal{O}(\delta^3)\right)$. Don't forget to reverse the substitution at the end of your calculation.

Question 4 ___/5

In epidemiology one tries to find out how much a certain exposure E (e.g. smoking) increases the frequency to get a certain disease D (e.g. cancer). The *relative risk*

$$RR\,(D) = \frac{P\,(D \mid E)}{P\,(D \mid \overline{E})}$$

is of central interest. Epidemiologists use further the odds

$$O(D \mid E) = \frac{P(D \mid E)}{P(\overline{D} \mid E)} \text{ as well as } O(D \mid \overline{E}) = \frac{P(D \mid \overline{E})}{P(\overline{D} \mid \overline{E})}$$

and the odds ratio (OR)

$$OR(D \mid E) = \frac{O(D \mid E)}{O(D \mid \overline{E})} .$$

What are possible implications of those quantities and what do they tell us?

a) In media reports about epidemiological studies often the relative risk is (indirectly) mentioned, although the study actually estimated the odds ratio. Which condition has to be satisfied that the relative risk is equal to the odds ratio? Argue why this condition is approximately fulfilled in many realistic situations. ___/2

b) Obtaining $P(D \mid E)$ and $P(D \mid \overline{E})$ to calculate the relative risk, via a so called cohort study is very expensive as the rareness of diseases typically requires that many people with E and $\overline{E}$ are monitored. Therefore, much cheaper 'case control studies' are performed usually, in which sick (D) and non-sick ($\overline{D}$) people are interviewed about their exposure. This provides $P(E \mid D)$ and $P(E \mid \overline{D})$. Show that this is sufficient to calculate the odds ratios $OR(E \mid D)$ and $OR(D \mid E)$, where the latter is often – to a good approximation – also the relative risk. ___/3

Question 5 ___/19

A sports game in which two players play for a point in each round, the game is won by the first player who leads by two points. Player A always has the probability θ to win a round. What are his chances?

a) State the three probabilities (as a function of θ) that after two rounds player has A won, has lost, or has to continue the game. Call them in the following w, l, and c, respectively. ___/3

b) State now the three probabilities (as a function of w, l, and c) that after (at most) four rounds player A has won, has lost, or that the game still continues. ___/4

c) State now the three probabilities that after (at most) an infinite numbers of rounds player A has won, lost, or that the game still continues. ___/3

d) Now you hear that player A has won the game. What is the probability $P(n|W)$ (as a function of θ) that the game ended with round n? If you were not able to solve c) use the following: ___/3

$$P(W|n \leq \infty) = \frac{\theta^2}{1 - 2\theta(1-\theta)} \qquad P(L|n \leq \infty) = \frac{(1-\theta)^2}{1 - 2\theta(1-\theta)} \qquad P(C|n \leq \infty) = 0$$

e) How much information (in bits) on the length of the game did you get from hearing that A is the winner? ___/2

f) How much information on the winner of the game do you get from hearing that player A has won, depending on your a priori knowledge about θ? Examine three interesting cases, $\theta = 1$, $\theta = 1/2$ and $\theta \ll 1$ (expand to leading order). ___/4

Question 6 ___/A

ssume that for a PDF $P(x)$ only a set of constraints —typically moment constraints— are known, like

$$
\begin{aligned}
Q\left(x\right)=P\left(x\mid I\right)&=\text{const}, && \text{(background information)}, && \text{(1a)}\\
P(x)&>0, && \text{(positive definitness)}, && \text{(1b)}\\
\int P(x)\,\mathrm{d}x&=1, && \text{(normalization)}, && \text{(1c)}\\
\int P(x)f_i(x)\,\mathrm{d}x&=\alpha_i, && \text{(moment constraints)}. && \text{(1d)}
\end{aligned}
$$

From the principle of maximum entropy one derives that

$$P(x)=\exp\left(\lambda_0+\sum_{i=1}^{n}\lambda_i f_i(x)\right)=\frac{\exp\left(\sum_{i=1}^{n}\lambda_i f_i(x)\right)}{Z}=\frac{Q\left(x\right)}{Z\left(\mu\right)}e^{\mu f(x)} \tag{2}$$

with λ_i such that the constraints are satisfied.

a) Assume a PDF $P\left(x\mid\lambda\right)$ with $x\in\mathbb{R}$. Show that no maximum entropy function exist if the only non-trivial constraint solely affects the first moment, $\langle x\rangle_{(x|\lambda)}=\lambda$ (1 point). __/

b) Now take the above case and add the restriction that the support of x is only on $\mathbb{R}^+$. Derive the corresponding maximum entropy PDF (2 points). __/

c) Now assume that $\langle\ln x\rangle_{(x|\alpha)}=\alpha$, with $\alpha>0$ and $x\in[1,\infty)$ is known. Calculate the PDF (2 points). __/

Question 7

__/6

You are interested in two numbers, $s=(s_1,s_2)\in\mathbb{R}^2$. Your measurement device, however, only measures sum and difference of the numbers, according to

$$
\begin{aligned}
d_1&=s_1+s_2+n_1\\
d_2&=s_1-s_2+n_2
\end{aligned}
$$

with some noise vector $n\in\mathbb{R}^2$. Assume a Gaussian prior $\mathcal{P}(s)=\mathcal{G}(s,S)$ for s and a Gaussian PDF for the noise, $\mathcal{P}(n)=\mathcal{G}(n,N)$, with $N_{ij}=\sigma^2\delta_{ij}$.

a) Assume a flat prior. Write down the response matrix, and calculate the posterior $\mathcal{P}(s|d)$. Reminder: $\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}^{-1}=\frac{1}{\alpha\delta-\beta\gamma}\begin{pmatrix}\delta & -\beta\\ -\gamma & \alpha\end{pmatrix}$ __/3

b) In the setting from above, now assume a generic response $R=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}$. Calculate the mean of $\mathcal{P}(s|d)$. Discuss in which circumstances (regarding R) the linear filter fails and give the underlying reason. __/3

Question 8

__/6

Assume the measurement of a signal s, which yields the data d, leads to.

$$
\begin{aligned}
\mathcal{P}(s)&=\mathcal{G}(s,S)\\
\mathcal{P}(s|d)&=\mathcal{G}(s-m,D)
\end{aligned}
$$

a) Calculate the amount of information in terms of entropy one gains via the measurement. ___/4

b) Now assume that your signal prior in the above measurement was flat. How much information in terms of entropy does one gain via the measurement? Explain the result. ___/2

Scratch paper

Scratch paper

Scratch paper

Scratch paper

Scratch paper

Scratch paper