# Final Exam Probability Theory (WBMA046-05)

23 June 2023, 08:30-10:30

READ THE INSTRUCTIONS BELOW CAREFULLY BEFORE STARTING THE EXAM.

- This exam contains 8 pages (including this cover page) and 5 exercises.
- Write your name and student number at the top of EACH page (including this cover page).
- Your answers should be written in this booklet. Preferably, avoid handing in extra paper.
  If you do not have enough space to write your solution below the exercise, you can use the back of the sheet. If this is still not enough space, you can add one (or more) extra sheet(s) to the exam. In this case, indicate it clearly in the booklet, and write your name and student number on the top of that (these) sheet(s).
- Do not write on the table below.
- Sticking to the rules above is worth 10 points.
- It is absolutely NOT allowed to use calculator, phone, smartwatch, books, lecture notes or any other aids.
- Always give a short proof of your answer or a calculation to justify it, or clearly state the facts from the lecture notes you are using (unless it is stated explicitly in the question this is not needed.).

| Exercise | Points | Score |
|---|---|---|
| 1 | 22 | |
| 2 | 23 | |
| 3 | 15 | |
| 4 | 10 | |
| 5 | 20 | |
| Follow the rules (name on each page...) | 10 | |
| Total | 100 | |

| z | 0 | 0,01 | **0,02** | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,5 | 0,50399 | 0,50798 | 0,51197 | 0,51595 | 0,51994 | 0,52392 | 0,5279 | 0,53188 | 0,53586 |
| 0,1 | 0,53983 | 0,5438 | 0,54776 | 0,55172 | 0,55567 | 0,55962 | 0,56356 | 0,56749 | 0,57142 | 0,57535 |
| 0,2 | 0,57926 | 0,58317 | 0,58706 | 0,59095 | 0,59483 | 0,59871 | 0,60257 | 0,60642 | 0,61026 | 0,61409 |
| 0,3 | 0,61791 | 0,62172 | 0,62552 | 0,6293 | 0,63307 | 0,63683 | 0,64058 | 0,64431 | 0,64803 | 0,65173 |
| 0,4 | 0,65542 | 0,6591 | 0,66276 | 0,6664 | 0,67003 | 0,67364 | 0,67724 | 0,68082 | 0,68439 | 0,68793 |
| 0,5 | 0,69146 | 0,69497 | 0,69847 | 0,70194 | 0,7054 | 0,70884 | 0,71226 | 0,71566 | 0,71904 | 0,7224 |
| 0,6 | 0,72575 | 0,72907 | 0,73237 | 0,73565 | 0,73891 | 0,74215 | 0,74537 | 0,74857 | 0,75175 | 0,7549 |
| 0,7 | 0,75804 | 0,76115 | 0,76424 | 0,7673 | 0,77035 | 0,77337 | 0,77637 | 0,77935 | 0,7823 | 0,78524 |
| 0,8 | 0,78814 | 0,79103 | 0,79389 | 0,79673 | 0,79955 | 0,80234 | 0,80511 | 0,80785 | 0,81057 | 0,81327 |
| 0,9 | 0,81594 | 0,81859 | 0,82121 | 0,82381 | 0,82639 | 0,82894 | 0,83147 | 0,83398 | 0,83646 | 0,83891 |
| 1 | 0,84134 | 0,84375 | 0,84614 | 0,84849 | 0,85083 | 0,85314 | 0,85543 | 0,85769 | 0,85993 | 0,86214 |
| 1,1 | 0,86433 | 0,8665 | 0,86864 | 0,87076 | 0,87286 | 0,87493 | 0,87698 | 0,879 | 0,881 | 0,88298 |
| 1,2 | 0,88493 | 0,88686 | 0,88877 | 0,89065 | 0,89251 | 0,89435 | 0,89617 | 0,89796 | 0,89973 | 0,90147 |
| 1,3 | 0,9032 | 0,9049 | 0,90658 | 0,90824 | 0,90988 | 0,91149 | 0,91309 | 0,91466 | 0,91621 | 0,91774 |
| **1,4** | 0,91924 | 0,92073 | **0,9222** | 0,92364 | 0,92507 | 0,92647 | 0,92785 | 0,92922 | 0,93056 | 0,93189 |
| 1,5 | 0,93319 | 0,93448 | 0,93574 | 0,93699 | 0,93822 | 0,93943 | 0,94062 | 0,94179 | 0,94295 | 0,94408 |
| 1,6 | 0,9452 | 0,9463 | 0,94738 | 0,94845 | 0,9495 | 0,95053 | 0,95154 | 0,95254 | 0,95352 | 0,95449 |
| 1,7 | 0,95543 | 0,95637 | 0,95728 | 0,95818 | 0,95907 | 0,95994 | 0,9608 | 0,96164 | 0,96246 | 0,96327 |
| 1,8 | 0,96407 | 0,96485 | 0,96562 | 0,96638 | 0,96712 | 0,96784 | 0,96856 | 0,96926 | 0,96995 | 0,97062 |
| 1,9 | 0,97128 | 0,97193 | 0,97257 | 0,9732 | 0,97381 | 0,97441 | 0,975 | 0,97558 | 0,97615 | 0,9767 |
| 2 | 0,97725 | 0,97778 | 0,97831 | 0,97882 | 0,97932 | 0,97982 | 0,9803 | 0,98077 | 0,98124 | 0,98169 |
| 2,1 | 0,98214 | 0,98257 | 0,983 | 0,98341 | 0,98382 | 0,98422 | 0,98461 | 0,985 | 0,98537 | 0,98574 |
| 2,2 | 0,9861 | 0,98645 | 0,98679 | 0,98713 | 0,98745 | 0,98778 | 0,98809 | 0,9884 | 0,9887 | 0,98899 |
| 2,3 | 0,98928 | 0,98956 | 0,98983 | 0,9901 | 0,99036 | 0,99061 | 0,99086 | 0,99111 | 0,99134 | 0,99158 |
| 2,4 | 0,9918 | 0,99202 | 0,99224 | 0,99245 | 0,99266 | 0,99286 | 0,99305 | 0,99324 | 0,99343 | 0,99361 |
| 2,5 | 0,99379 | 0,99396 | 0,99413 | 0,9943 | 0,99446 | 0,99461 | 0,99477 | 0,99492 | 0,99506 | 0,9952 |
| 2,6 | 0,99534 | 0,99547 | 0,9956 | 0,99573 | 0,99585 | 0,99598 | 0,99609 | 0,99621 | 0,99632 | 0,99643 |
| 2,7 | 0,99653 | 0,99664 | 0,99674 | 0,99683 | 0,99693 | 0,99702 | 0,99711 | 0,9972 | 0,99728 | 0,99736 |
| 2,8 | 0,99744 | 0,99752 | 0,9976 | 0,99767 | 0,99774 | 0,99781 | 0,99788 | 0,99795 | 0,99801 | 0,99807 |
| 2,9 | 0,99813 | 0,99819 | 0,99825 | 0,99831 | 0,99836 | 0,99841 | 0,99846 | 0,99851 | 0,99856 | 0,99861 |
| 3 | 0,99865 | 0,99869 | 0,99874 | 0,99878 | 0,99882 | 0,99886 | 0,99889 | 0,99893 | 0,99896 | 0,999 |
| 3,1 | 0,99903 | 0,99906 | 0,9991 | 0,99913 | 0,99916 | 0,99918 | 0,99921 | 0,99924 | 0,99926 | 0,99929 |
| 3,2 | 0,99931 | 0,99934 | 0,99936 | 0,99938 | 0,9994 | 0,99942 | 0,99944 | 0,99946 | 0,99948 | 0,9995 |
| 3,3 | 0,99952 | 0,99953 | 0,99955 | 0,99957 | 0,99958 | 0,9996 | 0,99961 | 0,99962 | 0,99964 | 0,99965 |
| 3,4 | 0,99966 | 0,99968 | 0,99969 | 0,9997 | 0,99971 | 0,99972 | 0,99973 | 0,99974 | 0,99975 | 0,99976 |
| 3,5 | 0,99977 | 0,99978 | 0,99978 | 0,99979 | 0,9998 | 0,99981 | 0,99981 | 0,99982 | 0,99983 | 0,99983 |
| 3,6 | 0,99984 | 0,99985 | 0,99985 | 0,99986 | 0,99986 | 0,99987 | 0,99987 | 0,99988 | 0,99988 | 0,99989 |
| 3,7 | 0,99989 | 0,9999 | 0,9999 | 0,9999 | 0,99991 | 0,99991 | 0,99992 | 0,99992 | 0,99992 | 0,99992 |
| 3,8 | 0,99993 | 0,99993 | 0,99993 | 0,99994 | 0,99994 | 0,99994 | 0,99994 | 0,99995 | 0,99995 | 0,99995 |
| 3,9 | 0,99995 | 0,99995 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99996 | 0,99997 | 0,99997 |
| 4 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99997 | 0,99998 | 0,99998 | 0,99998 | 0,99998 |

Table 1: CDF $\Phi$ of a standard normal distribution.
Example: the value in the row “1,4” and column “0,02” gives the approximation $\Phi(1.42) \simeq 0.9222$.

**Exercise 1 (a:4, b:4, c:4, d:4, e:6 pts)** A group of 6 kids enters a room in which there are 3 toys. Each toy is picked by one kid. A kid might take more than one toy. If a kid does not get a toy they pick a book and read. For each of the following situations count how many ways there are to distribute the toys.
*Remark 1:* In any of the situations below the kids are distinguishable.
*Remark 2:* Give your answer (a number) directly in the box below. You are not required to provide any explanation in this exercise. For each question you get either zero or the full points.

1. The toys are distinct from one another and no kid picks more than one toy.

   The number of possibilities is:

2. The toys are distinct from one another. Any of the kids might pick several toys.

   The number of possibilities is:

3. The toys are identical and no kid picks more than one toy.

   The number of possibilities is:

4. The toys are identical. One kid picks 2 toys and one other picks 1 toy.

   The number of possibilities is:

5. The toys are identical. Any of the kids might pick several toys.

   The number of possibilities is:

**Solution**:

1. $(6)_3 = 6 \times 5 \times 4 = 120$
2. $6^3 = 216$
3. $\binom{6}{3} = \frac{6\times5\times4}{3\times2\times1} = 20$
4. $(6)_2 = 6 \times 5 = 30$
5. $20 + 30 + 6 = 56$ (sum the numbers of possibilities if 3 kids, 2 kids or only 1 kid get the toys)

**Exercise 2 (a:4, b:9, c:5, d:2, e:3 pts)**
The joint density of $X$ and $Y$ is given by

$$f(x,y) = C(y-x)e^{-y}\mathbf{1}(-y < x < y).$$

a) Find the value of the constant $C$.

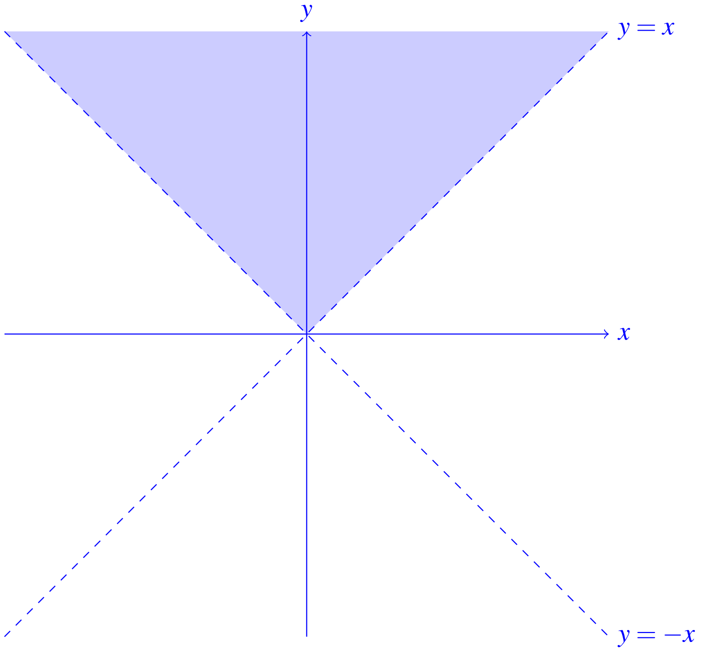
b) Find the density function $f_X$ of $X$.

c) Find $\mathbb{E}[X]$.

d) Let $f_Y$ denote the density function of $Y$. Show that $f_Y(y) > 0$ for any $y > 0$.

e) Determine if $X$ and $Y$ are independent.

*Hint:* You can use that $\int_0^\infty e^{-x}x^n\,\mathrm{d}x = n!$ for any $n \in \mathbb{N}$.
**Solution**: *Remark*: This exercise is the one that caused most troubles for students. Drawing the domain $\{(x,y) : -y < x < y\}$ helps a lot to avoid many of the mistakes that the lecturer and TAs read.




a) $1 = \int f(x,y)\,\mathrm{d}x\,\mathrm{d}y = C\int_0^\infty e^{-y}\int_{-y}^{y}(y-x)\,\mathrm{d}x\,\mathrm{d}y = C\int_0^\infty e^{-y}2y^2\,\mathrm{d}y = 4C$. Hence, $\boxed{C = 1/4}$.

b) For $x > 0$:

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,\mathrm{d}y = \int_x^\infty C(y-x)e^{-y}\,\mathrm{d}y = C\int_0^\infty ue^{-(x+u)}\,\mathrm{d}u = Ce^{-x}\int_0^\infty ue^{-u}\,\mathrm{d}u = Ce^{-x}\times(1!) = \boxed{\frac{1}{4}e^{-x}}.$$

For $x < 0$:

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,\mathrm{d}y = \int_{-x}^\infty C(y-x)e^{-y}\,\mathrm{d}y = C\left[-ye^{-y} - e^{-y} + xe^{-y}\right]_{-x}^\infty = \boxed{\frac{1}{4}(-2xe^x + e^x)}.$$

c)

$$\begin{aligned}
\mathbb{E}[X] &= \int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x \\
&= \frac{1}{4}\left[\int_{-\infty}^{0} (-2x^2e^x + xe^x)\,\mathrm{d}x + \int_0^{\infty} xe^{-x}\,\mathrm{d}x\right] \\
&= \frac{1}{4}\left[-\int_0^{\infty} (2u^2e^{-u} + ue^{-u})\,\mathrm{d}u + 1\right] \\
&= \frac{1}{4}\left[-(2\times(2!)+1!)+1\right] \\
&= \boxed{-1}.
\end{aligned}$$

d) For $y>0$, $f_Y(y) = \int_{-y}^{y} C(y-x)e^{-y}dx = Ce^{-y}\int_0^{2y} t\,dt > 0$.

e) They are $\boxed{\text{not independent}}$, because $f \neq f_X \times f_Y$ as we can see with $(x,y)=(2,1)$ for example. Indeed $f(2,1) = 0 \neq f_X(2)\times f_Y(1)$ since $f_X(2) = \frac{1}{4}e^{-2}$ and $f_Y(1) = \int_{-1}^{1} C(1-x)e^{-1}dx > 0$.

**Exercise 3 (a:5, b:10 pts)**

A continuous random variable $X$ has *Pareto distribution* with positive parameters $a$ and $\lambda$ if its pdf is

$$f(x) = \mathbf{1}(x > a)\lambda a^{\lambda} x^{-\lambda-1}.$$

a) Compute the cdf $F_X$ of $X$.

b) Find the distribution of $Z = \log(X/a)$.
*(Compute either its cdf or its pdf. If you recognize a classical distribution, name it.)*

**Solution**:

a) $\boxed{\text{For } x \leq a,\ F_X(x) = 0}$. For $x > a$:

$$F_X(x) = \int_a^x \lambda a^{\lambda} t^{-\lambda-1} \mathrm{d}t = \left[-a^{\lambda} t^{-\lambda}\right]_{t=a}^{x} = \boxed{1 - a^{\lambda} x^{-\lambda}}.$$

b) Let $h\colon [1,\infty) \to [0,\infty)$ defined by $h(x) = \log(x/a)$. We have $Z = h(X)$ and $h$ is a strictly increasing differentiable function with inverse $h^{-1}(z) = ae^z$. In particular $Z$ takes value in $[0,\infty)$ (the codomain of $h$), and for $z \in [0,\infty)$ we have

$$f_Z(z) = f_X(h^{-1}(z)) \left| \frac{\partial h^{-1}(z)}{\partial z} \right| = \lambda a^{\lambda} (ae^z)^{-\lambda-1} a e^z = \lambda e^{-\lambda z}.$$

Thus we find that $\boxed{Z \text{ has an exponential distribution of rate } \lambda.}$

**Exercise 4 (10 pts)**

Assume $Y, X_1, X_2, \dots$ are independent and discrete, $Y$ takes values in $\{0,1,2,\dots\}$ and $X_1, X_2, \dots$ are identically distributed. Let

$$Z = \begin{cases} 0 & \text{if } Y = 0; \\ X_1 + \cdots + X_Y & \text{if } Y > 0. \end{cases}$$

Show that

$$\mathbb{E}(Z) = \mathbb{E}(Y) \cdot \mathbb{E}(X_1).$$

**Solution**:

$$\begin{aligned}
\mathbb{E}[Z] &= \mathbb{E}\left[\sum_{i=1}^{\infty} \mathbf{1}(Y=n)(X_1+\cdots+X_n)\right] \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbf{1}(Y=n)(X_1+\cdots+X_n)\right] \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbf{1}(Y=n)\right] \times \mathbb{E}\left[X_1+\cdots+X_n\right] && \text{because } Y \text{ and } X_1+\cdots+X_n \text{ are independent} \\
&= \mathbb{E}\left[X_1\right] \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbf{1}(Y=n)\right] \times n && \text{because } X_1,\dots,X_n \text{ are identically distributed} \\
&= \mathbb{E}\left[X_1\right] \sum_{i=1}^{\infty} \mathbb{P}\left[Y=n\right] \times n \\
&= \mathbb{E}\left[X_1\right] \mathbb{E}\left[Y\right].
\end{aligned}$$

**Exercise 5 (a:10, b:10 pts)**

An astronomer is interested in measuring the distance, in light-years, from his observatory to a distant star. Although the astronomer has a measuring technique, he knows that because of changing atmospheric conditions and normal error, each time a measurement is made, it will not yield the exact distance, but merely an estimate. As a result, the astronomer plans to make a series of measurements and then use the average value of these measurements as his estimated value of the actual distance.

We assume that the values of the measurements are independent and identically distributed random variables having a common mean $d$ (the actual distance) and a common variance of 4 (light-years).

On one hand the astronomer wants to take sufficiently many measurements to be reasonably sure that his estimated distance is accurate to within .5 light-years. On the other hand, booking a telescope to make his observations is very expensive so he does not want to do many more observations than needed.

a) Let $n$ be the number of measurements made. And let $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$ denote the average of the measurements. Show that the probability of the event $\{\overline{X_n} \in [d-0.5, d+0.5]\}$ can reasonably be approximated by $2\Phi(\sqrt{n}/4) - 1$, where $\Phi$ denotes the cdf of the standard normal distribution.

b) The astronomer can take 45 measurements per night. How many nights does he need to book the telescope if he wants to be 95% sure that $\overline{X_n} \in [d-0.5, d+0.5]$?
*Remark:* For this question, we assume that the approximation of the previous question is very good.

**Solution**:

a)

$$\mathbb{P}\left(\overline{X_n} \in [d-0.5, d+0.5]\right) = \mathbb{P}\left(-0.5 \times \frac{\sqrt{n}}{2} \leq \frac{\sum_{i=1}^n X_i - nd}{\sqrt{4 \times n}} \leq 0.5 \times \frac{\sqrt{n}}{2}\right)$$

Since $\mathbb{E}X_i = d$ and $\mathbb{V}\mathrm{ar}X_i = 4$, from the central limit theorem, we have that

$$\begin{aligned}
\mathbb{P}\left(\overline{X_n} \in [d-0.5, d+0.5]\right) &\simeq \mathbb{P}\left(-0.5 \times \frac{\sqrt{n}}{2} \leq Z \leq 0.5 \times \frac{\sqrt{n}}{2}\right) \qquad \text{(where } Z \sim \mathcal{N}(0,1)\text{)} \\
&= \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) \\
&= \Phi\left(\frac{\sqrt{n}}{4}\right) - \left(1 - \Phi\left(\frac{\sqrt{n}}{4}\right)\right) \\
&= 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1.
\end{aligned}$$

b) The number $n$ of measurements needs to satisfy $2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \geq 0.95$.
This is equivalent with $\Phi\left(\frac{\sqrt{n}}{4}\right) \geq (1+0.95)/2 = 0.975 = \Phi(1.96)$.
Thus we want to have $\frac{\sqrt{n}}{4} \geq 1.96$, or equivalently $n \geq (4 \times 1.96)^2 = (7.84)^2$.
In particular $n > 7^2 = 49 > 45$ and $n < 8^2 = 64 < 2 \times 45$.
That means that one night of observation is not sufficient, but two nights is.
Hence $\boxed{\text{the astronomer needs to book the telescope 2 nights.}}$