

# TENTAMEN I MATEMATISK STATISTIK FÖR TILLÄMPAD ARTIFICIELL INTELLIGENS

7.5 HP

januari, 2024

**Maxpoäng:** 30p.    **Betygsgränser:** 12p: betyg 3, 18p: betyg 4, 24p: betyg 5.

**Hjälpmedel:** Valfri miniräknare samt formelsamling som medföljer tentamenstexten.

**Kursansvarig:** Eric Järpe, telefon 0729-77 36 26.

Till uppgifterna skall *fullständiga lösningar* lämnas. Lösningarna ska vara *utförligt* redovisade! Bladen ska lämnas in i rätt ordning. Svara alltid med 4 decimalers noggrannhet om ej annat anges. Lösningar kommer finnas på internet: <http://dixon.hh.se/erja/teach> → Matematisk statistik.

1. Låt funktionen  $F$  vara definierad av

$$F(x) = \frac{1}{2} \left( 1 + \sin \left( \frac{\pi x}{2(\theta + |x|)} \right) \right)$$

för alla  $x \in \mathbb{R}$  och  $\theta \in \mathbb{R}^+$ .

- (a) Verifiera att  $F$  är en fördelningsfunktion för en slumpvariabel. (3p)
- (b) Beräkna  $P(X \leq 2)$  exakt om  $\theta = 1$ . (3p)
- (c) Antag att slumpvariabeln  $X$  har  $F$  som fördelningsfunktion. Beräkna täthetsfunktionen  $f$  för  $X$  och bevisa att

$$\lim_{x \rightarrow 0-} f(x) = \lim_{x \rightarrow 0+} f(x)$$

för alla  $\theta \in \mathbb{R}^+$ . (4p)

2. Vid exekvering av 7 körningar med en viss AI-arkitektur är tider och motsvarande datamängder

<i>Exekveringstid (minuter)</i>	321	182	117	399	75	210	154
<i>Datamängd (GB)</i>	103.2	25.5	22.8	99.7	13.0	72.1	50.5

- (a) Beräkna den tredje kvartilen av variabeln *Exekveringstid* baserat på dessa AI:ar. (2p)
- (b) Vad blir den bäst passande linjära modellen av dessa data med *Datamängd* som responsvariabel och *Exekveringstid* som kovariat? (3p)

3. Antag att man använder en ny AI för bedömning av ögonkontroller med hjälp av en ögonbottenscanning hos diabetiker. En skillnad mot mänsklig bedömning är den avsevärt mycket snabbare svarstiden. Man vill dock även att noggrannheten vid bedömningen ska vara hög.

- (a) I en grupp med 30 diabetiker har exakt 25 stycken någon form av problem som behöver upptäckas vid kontrollen. Vad är sannolikheten att minst 95% av 20 patienter, som är slumpvis valda bland de 30, har dessa problem? (3p)
- (b) Vad blir styrkan av ett hypotestest med 50 patienter av om AI:n klassificerar mer än 95% korrekt om den i själva verket klassificerar exakt 97% korrekt? (4p)
- (c) Antag att man scannar 37 patienter och att AI:n klassificerar 35 patienter korrekt. Gör ett hypotestest på 5% signifikansnivå av om sannolikheten att AI:n bedömer en patient korrekt är större än 90%. Vad blir  $p$ -värdet? (3p)

En viss form av problem brukar, med rätt behandling, försvinna efter ungefär två veckor. Om man låter  $X$  vara tiden efter diagnos som detta problem fortfarande är kvar så är  $X$ , räknat i dagar, en exponentialfördelad slumpvariabel med parameter  $\lambda = 0.07$ . Beräkna

- (d) sannolikheten att en person fortfarande har besvär efter 17 dagar? (2p)
- (e) den betingade sannolikheten att en person har besvär efter 17 dagar givet att hen har det efter 10 dagar? (3p)
- (f) den approximativa sannolikheten att det genomsnittliga antalet dagar som en person har kvar besvär är minst 14.5 dagar efter diagnosen, beräknat på en grupp med 10 personer? (3p)

*LYCKA TILL!*