Tentamentsskrivning i **Mathematisk statistik TMA321, 3p.**

Tid: Tisdagen den 22 augusti, 2006 kl 8.30-12.30
Examinator och jour: Serik Sagitov, tel. 772-5351, mob. 0736 907 613, MV-huset rum H3026.
Hjälpmedel: valfri räknare, egen formelsamling (4 sidor på 2 blad A4) samt utdelade tabeller.

There are five questions with the total number of marks 30. Attempt as many questions, or parts of the questions, as you can. Preliminary grading system (no bonus points this time):
grade "3" for 12 to 17 marks,
grade "4" for 18 to 23 marks,
grade "5" for 24 and more marks.

---

1. (6 marks) Figure 1 gives a discrete distribution of hits of penalty shots for a certain football player. Here the football goal is outlined by three double lines. According to these numbers the probability to miss the goal is 10%.

a. Let X and Y be the coordinates of a random point hit by a penalty shot. Judging from the given discrete distribution are X and Y independent? Explain.

b. Give the joint distribution of X and Y conditioned on the event that the random point is inside the frame of the goal.

c. Figure 2 contains a table with the conditional probabilities of successful penalties. For example, if the ball hits an upper corner of the goal, the probability to score is 0.9. (Notice that these probabilities do not sum up to 100%.) Using this table find the total probability of a penalty to score a goal.

2. (6 marks) The true distribution of the horizontal coordinate X for the penalty shot is continuous. Figure 3 presents a continuous model for the football goal of width 4 m. It reflects the following strategy: the player tosses a coin and chooses with equal probabilities one of the two normal distributions $N(1,\sigma^2)$ or $N(-1,\sigma^2)$ where $\sigma = 0.5$.

a. Find the probability of a miss for the continuous model. Explain every step in your calculations.

b. The continuous model implies the following representation

$$X = I \cdot X_1 + (1 - I) \cdot X_2, \quad (1)$$

where I, X_1 and X_2 are three independent random variables with the marginal

1	1	1	1	1	1
1	10	5	5	10	1
1	25	5	5	25	1

Figure 1: Percentages of penalty hits

0	0	0	0	0	0
0	90	60	60	90	0
0	70	40	40	70	0

Figure 2: Conditional probabilities (in percent) of a successful penalty.

distributions $X_1 \sim N(1, (0.5)^2)$, $X_2 \sim N(-1, (0.5)^2)$ and

$$P(I = 0) = P(I = 1) = 1/2.$$

Are the two components $I \cdot X_1$ and $(1 - I) \cdot X_2$ independent? Explain. Find the mean of X from (1) using the general properties of the mean value of a sum and a product.

c. Explain the properties of the indicator random variable I

$$I^2 = I,\ (1 - I)^2 = 1 - I,\ I(1 - I) = 0, \tag{2}$$

and then find the standard deviation of X. Hint: compute $E(X^2)$ using (1) and (2).

3. (6 marks) Opponents of the construction of a dam on the New River claim that less than half of the residents living along the river are in favor of its construction. A survey is conducted to gain support for this point of view.

a. Set up the appropriate null and alternative hypotheses.

b. Find the rejection region for the null hypothesis on 10% significance level.

c. Of 500 people surveyed, 230 favor the construction. Is this sufficient evidence to justify the claim of the opponents of the dam?

d. To what type of error are you now subject? Discuss the practical consequences of making such an error.

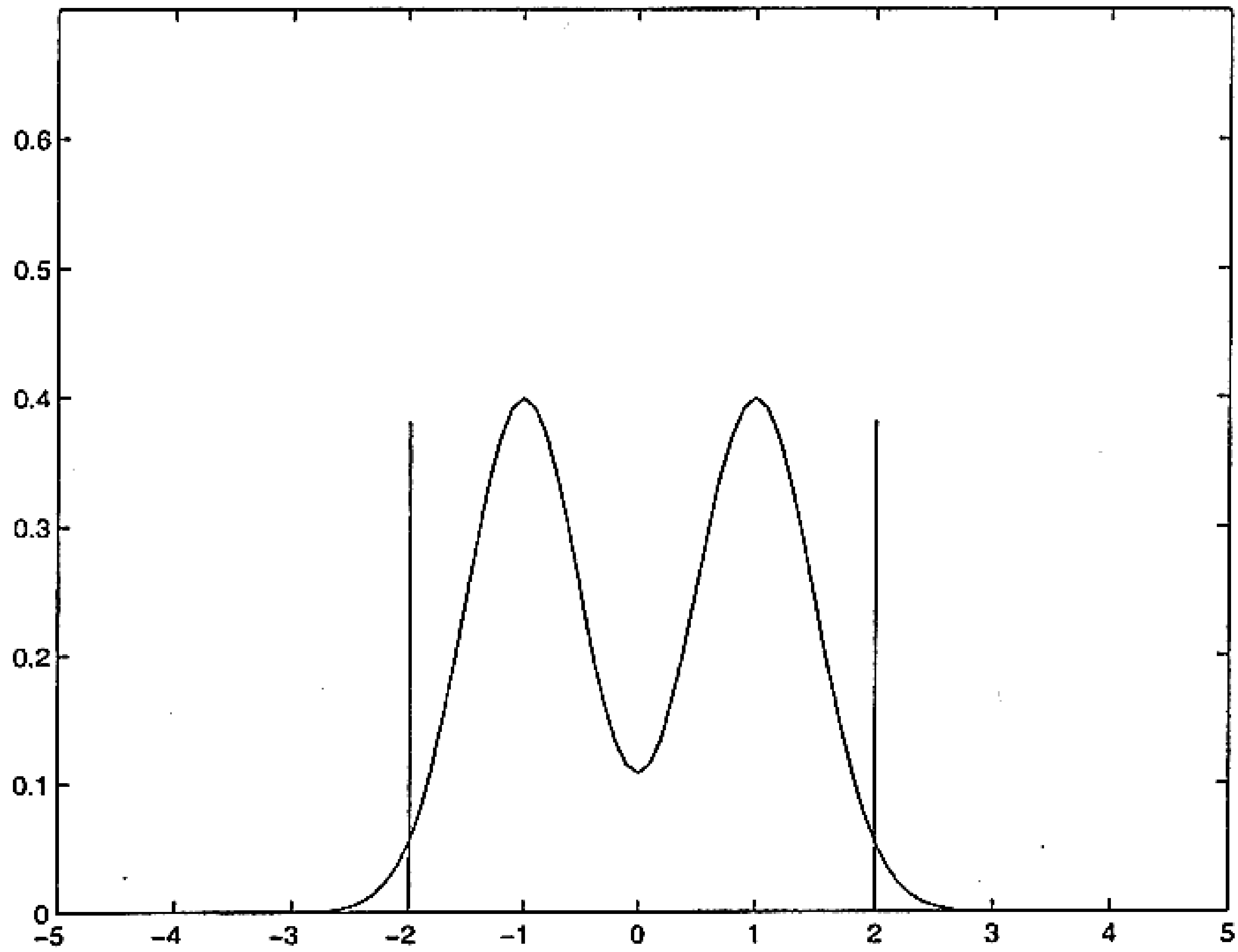

Figure 3: *Continuous X*

4. (6 marks) Computer terminals have a battery pack that maintains the configuration of the terminal. These packs must be replaced occasionally. Let X denote the life span in years of a such a battery. Assume that X has probability density function $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. The following data on the X values has been collected

1.7	4.0	1.9	2.0	1.7
2.1	2.7	4.2	1.8	2.2
3.1	1.5	2.4	6.2	7.0
3.6	1.4	5.0	3.8	1.6

a. Plot the data to see if it really fits the exponential distribution $f(x) = \lambda e^{-\lambda x}$.

b. Find a method of moment estimate for the parameter λ.

c. Compute an approximate 95% CI for λ. Hint: first find a CI for the mean $\mu = E(X)$.

5. (6 marks) A simple linear regression model predicting the height (in cm) of a daughter Y from the height of the mother X has a linear form $Y = 84 + 0.5X + \epsilon$, where ϵ is normally distributed with zero mean and standard deviation 5.8. The model was derived by means of the least square method from a data set of size $n = 45$ collected from a certain population.

a. What kind of factors does the term ϵ stand for? Are these factors supposed to be independent form the main factor X? How the normality assumption is usually justified? Give a detailed answer.

b. Judging from the model sketch a scatterplot indicating how the data might look like. Use this plot to illustrate the "regression to mediocrity" phenomenon.

c. The variation in mothers' and daughters' heights was found to be nearly equal. What percentage in daughters' height variation is explained by the mothers' height variation?

Statistical tables supplied:

1. Normal distribution table
2. t-distribution table

Good luck!