STAT 391  
Final Exam Solutions  
June 10, 2002  
©Marina Meilä  
mmp@cs.washington.edu

**Problem 1**

Below are depicted three kernel density estimates of the same density. All of them use the same kernel function. Two different data sets $D_1$, $D_2$ and two different kernel widths $h_1$, $h_2$ have been used to make these plots.

a. What kernel was used to make the plots?

- [ ] Square  
- [ ] Epanechnikov  
- [ ] Gaussian

b. Which kernel width is larger?

- [ ] $h_1$  
- [x] $h_2$

c. Which data set is larger?

- [ ] $D_1$  
- [x] $D_2$

**Problem 2**

a. What is the normalization constant $Z$?

$$Z = 10$$

b. Compute the probability $P(X < 1)$.

$$P(X < 1) = \frac{1}{10}$$

c. Compute the probability $P(1.2 < X < 2)$

$$P(1.2 < X < 2) = \frac{16}{100}$$

e. Mark on the graph above the position of $x_{1/2}$, the median of $X$.

$$x_{1/2} = 2 + \frac{2}{3} = \frac{8}{3}$$
Problem 3

a. What is $P(Y < 1)$?

Let $Y' = Y - 3$, so that $Y' \sim N(0,1)$. Then, $P(Y < 1) = P(Y' < -2) = P(Y' > 2) = p_2$.

c. What is $P(Y < 3)$?

$$P(Y < 3) = P(Y' < 0) = \frac{1}{2}$$

d. What is $P(|Z| > 2)$?

Let $Z = 2Z'$ so that $Z' \sim N(0,1)$. Then

$$P(|Z| > 2) = P(Z > 2) + P(Z < -2) = 2P(Z' > 1) = 2p_1$$

e. What is Cov($Y, Z$)?

$$Cov(Y, Z) = \rho(Y, Z) \times \sqrt{\text{Var}(Y)\text{Var}(Z)}$$

$$= \frac{1}{\sqrt{2}} \cdot 1.2 = \sqrt{2}$$

f. Compute $P(Z < 0|Y = 4)$.

First we need to calculate the conditional distribution of $Z$ given $Y = 4$. This is normal distribution, whose mean and variance are

$$\mu_{Z|Y=4} = \mu_Z + (y - \mu_Z) \frac{\sigma_{YZ}}{\sigma_Y^2}$$

$$= 0 + (4 - 3) \frac{\sqrt{2}}{1}$$

$$= \sqrt{2}$$

$$\sigma^2_{Z|Y} = \sigma_Z^2(1 - \rho^2_{Y,Z})$$

$$= 2$$

Assume $Z'' \sim N(0,1)$ is a random variable such that

$$Z'' = \frac{Z - \mu_{Z|Y=4}}{\sigma_{Z|Y}}$$

Then

$$P(Z < 0|Y = 4) = P(Z'' < \frac{0 - \mu_{Z|Y=4}}{\sigma_{Z|Y}}) = P(Z'' < -1) = p_1$$
Figure 1: The areas enclosed by the thick lines are equal and represent equal probabilities in all four plots.

Problem 4

a. Estimate the distribution of $X$, the grade in Palmistry, in last year’s class.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$P(X)$</th>
<th>$A$</th>
<th>$B$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\frac{20}{20} = .6$</td>
<td>$\frac{12}{20} = 0.3$</td>
<td>$\frac{8}{20} = 0.1$</td>
</tr>
</tbody>
</table>

b. Estimate the probability of the event “$X = Y$”.

$$P(X = Y) = \frac{8}{20} = 0.4$$

c. Are the events “$X = Y$” and “$F$ in at least one test” independent?

Let $Q$ = “$F$ in at least one test”.

$$P(Q) = \frac{2}{20}$$

$$P(Q, X = Y) = \frac{1}{20}$$

$$P(Q)P(X = Y) = \frac{16}{400} = \frac{1}{25} \neq \frac{1}{20} = P(Q, X = Y)$$

The events are not independent.
d. To pass each test, one needs a grade of $A$ or $B$ in that test. What is the probability that a student passes Palmistry?

Let $R =$ “student passes Palmistry”.

$$P(R) = P(X = A \text{ or } X = B) = \frac{18}{20} = 0.9$$

e. What is the probability that a student passes Palmistry given that (s)he gets a $B$ in Tea Leaves?

$$P(X = A \text{ or } Y = B) = \frac{P(X = A \text{ or } Y = B)}{P(Y = B)} = \frac{3}{4} = 0.75$$

f. Are $X$ and $Y$ independent?

$$P(X = A, Y = A) = \frac{4}{20} = 0.2 \neq P(X = A)P(Y = A) = \frac{12}{20} \frac{7}{20} = 0.21$$

They are not independent.

g. This year’s class has $n' = 40$ students. What is probability that all students pass both tests?

$$P(\text{pass both tests}) =
= P(X = A, Y = A) + P(X = A, Y = B) + P(X = B, Y = A) + P(X = B, Y = B)
= \frac{4}{20} + \frac{8}{20} + \frac{3}{20} + \frac{3}{20}
= \frac{18}{20} = 0.9$$

The desired probability is $(0.9)^{40}$.

h. What is the expected number of students in this year’s class that pass both tests?

We have a Binomial random variable with $p = 0.9$ and $n = 40$. The mean value is $np = 36$.

i. Let $S =$”Snape assigned this grade”.

$$P(S|X = F) = \frac{P(X = F|S)P(S)}{P(X = F)}$$
\[
\begin{align*}
&= \frac{P(X = F|S)}{P(X = F|S) + P(X = F|S\,^c)P(S\,^c)} \\
&= \frac{0.6 \times 0.2}{0.6 \times 0.2 + 0.1 \times 0.8} \\
&= \frac{3}{5} = 0.6
\end{align*}
\]

**Problem 5**

Paragon Pictures is planning to release two movies this summer season: “Attack From Planet Entropy” and “Rainier Raging”. Let \( A \) be the event that “Attack” is a success. The probability of \( A \) is

\[ p_A = \frac{1}{10} \]

Let \( R \) be the event that “Rainier” is a success. The probability of \( R \) is

\[ p_R = \frac{1}{5} \]

The summer season lasts 100 days and the movies’ planned release dates are \( t_A = 20 \) and \( t_R = 60 \). A movie is a success or a failure independently of any other movie from the same studio.

What isn’t known yet at Paragon Pictures is that the romantic comedy “Doll story” produced by a small independent studio is going to be a huge success this season, shadowing everything else after it is released. The release time \( t \) of “Doll story” is unknown, and its density \( f_t \) is uniform in the interval \([0, 100] \).

The probability that “Attack” is a success in the new conditions is

\[
P(A|t) = \begin{cases} 
p_A \frac{t_A - t}{100} & \text{if } t \leq t_A \\
p_A & \text{if } t > t_A \end{cases}
\]

“Rainier” is luckier because

\[
P(R|t) = \begin{cases} 
p_R \frac{t_R - t}{30} & \text{if } t \leq t_R \\
p_R & \text{if } t > t_R \end{cases}
\]

**a.** Compute the probability that “Attack” is a success.

\[
P(A) = \int_0^{100} P(A|t)f_t(t)\,dt
\]

\[
= \int_0^{20} p_A \frac{t_A - t}{100} \,dt + \int_{t_A}^{100} p_A \frac{1}{100} \,dt
\]

\[
= \frac{p_A}{10000} \left( 20 \int_0^{20} (20 - t)\,dt + p_A \frac{100 - 20}{100} \right)
\]

\[
= 0.082
\]
b. Compute and plot the probability density of $t$, the release time of “Doll Story” given that “Attack” was a success.

$$f(t|A) = \frac{f_t(t)P(A|t)}{P(A)}$$

$$= \begin{cases} \frac{p_A}{P(A)} & \text{if } t > t_A \\ \frac{p_A}{P(A)} \frac{t_A - t}{t_A} & \text{if } t \leq t_A \end{cases}$$

![Figure 2: Conditional probability density function for $t$.](image)

c. At time $t_A$ “Doll story” was not released yet. Could Paragon Pictures increase the probability of success of “Rainier raging” by changing its release date? More precisely, compute the time $t^*_R$ that maximizes the probability of success of “Rainier raging”.

The conditional density of $t$ given that $t > t_A$ is

$$f^{new}_t(t) = \begin{cases} 0 & \text{if } t \leq t_A \\ \frac{1}{50} & \text{otherwise} \end{cases}$$

Now we write $P(R)$ as a function of $t_R$

$$P(R) = \int_0^{t_R} P(R|t)f^{new}_t(t) \, dt$$

$$= \int_{t_A}^{t_R} p_R \frac{t_R - t}{30} \frac{1}{80} \, dt + \int_{t_R}^{100} p_R \frac{1}{80} \, dt$$

$$= \frac{p_R}{80} \left[ \frac{(t_R - t_A)^2}{2 \times 30} + (100 - t_R) \right]$$
\[ t_R^2 - 2(t_A + 30)t_R + \ldots \]

The above is a quadratic expression in \( t_R \). It has a minimum at \( t_A + 30 = 50 \) which is inside the interval \([20, 100]\). Hence, the maximum of \( P(R) \) is reached at the extremity of the interval that is furthest away from the minimum. This is \( t_R = 100 \).

![Figure 3: The parabola \( t_R^2 - 2(t_A + 30)t_R \) (plus a constant) showing the variation of \( P(R) \) with \( t_R \).](image)