STAT 391
Final Exam Solutions
June 8, 2006

©Marina Meilă and Lei Xu
mmp@cs.washington.edu, xul@u.washington.edu

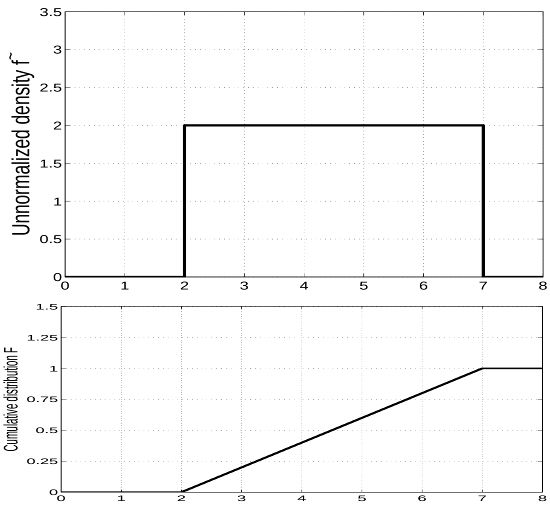
Problem 1 14 points

The following graphs represent unnormalized probability densities over the real axis. Assume that the densities are 0 outside the interval $[0, 8]$. Showing detail work is not required for this problem.

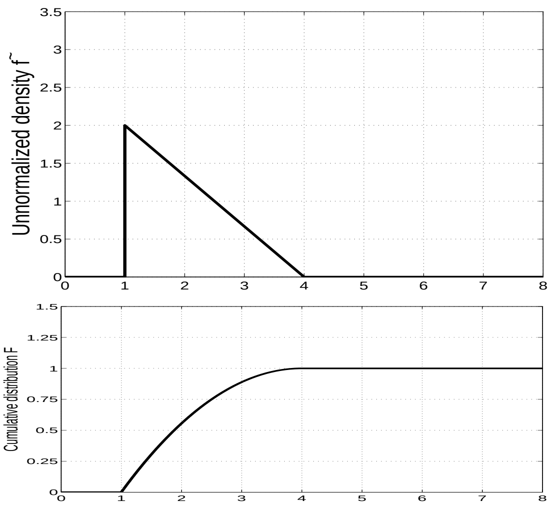
1.1 Compute the normalization constants for the four densities; write their values above the respective graphs. The normalization constant is the number Z so that $f = \frac{1}{Z}\tilde{f}$ is a probability distribution. 2 p.

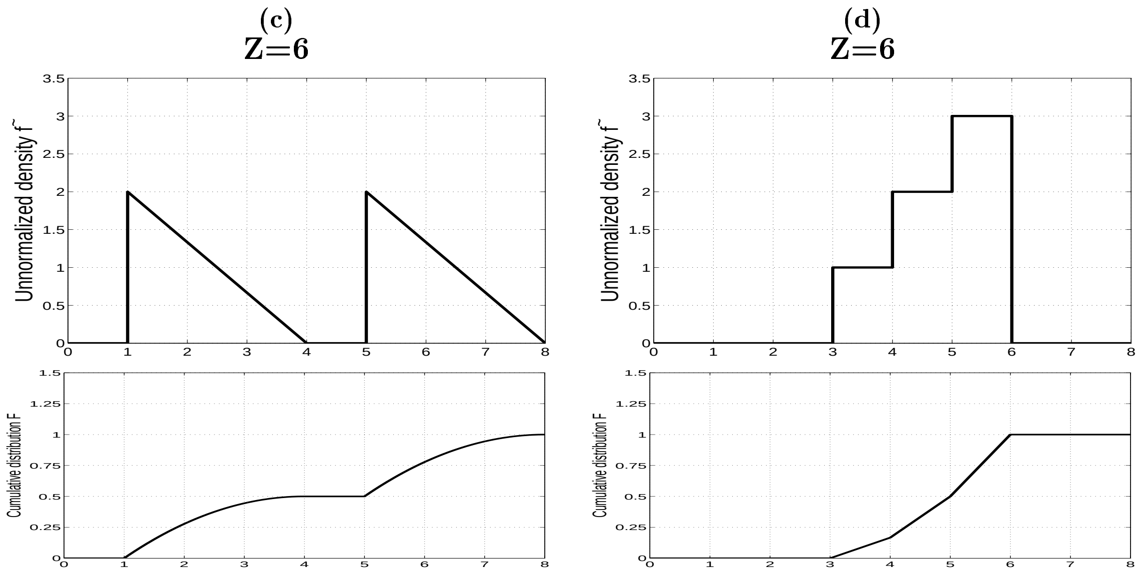
1.2 Draw the CDF's of the four densities in the corresponding plots. 4 p.

(a)
Z=10



(b)
Z=3





1.3 Compute the probability of the interval $[4, 6]$ under each of the four probability distributions. 2 p.

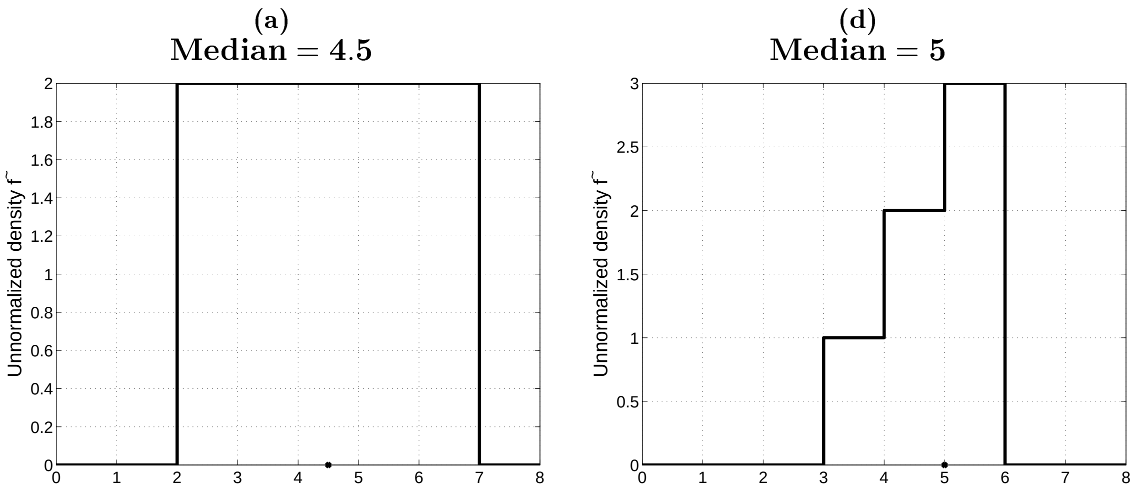
a. $P([4, 6]) = 2 * 2/10 = 0.4$

b. $P([4, 6]) = 0$

c. $P([4, 6]) = (2 + \frac{4}{3})/(2 \times 6) = 5/18$

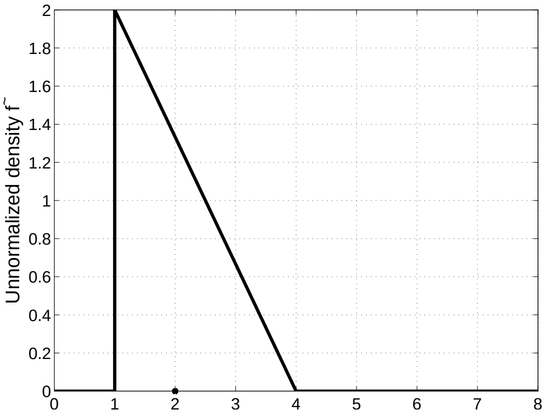
d. $P([4, 6]) = 1 - 1 * 1/6 = 5/6$

1.4 Mark on graphs **a, d** the location of the median. 1 p.

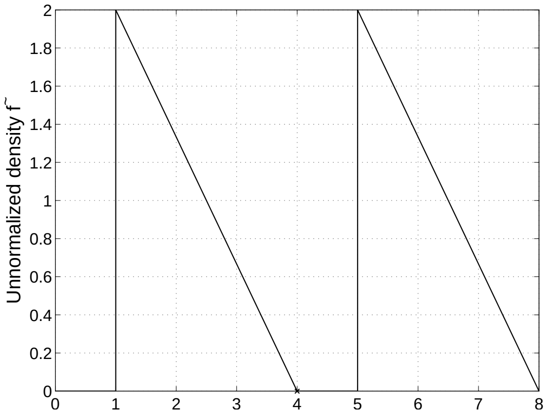


1.5 Mark on graphs **b, c** the location of the mean. 2 p.

(b)
Mean = 2



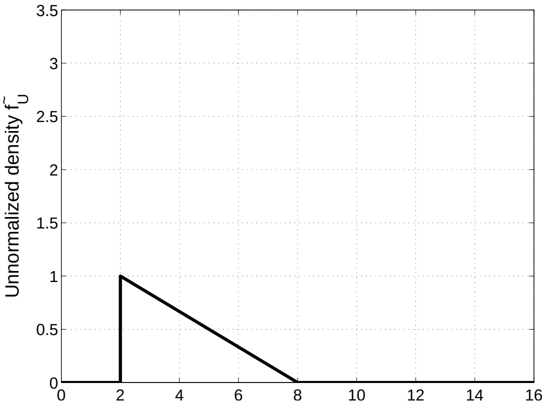
(c)
Mean = 4



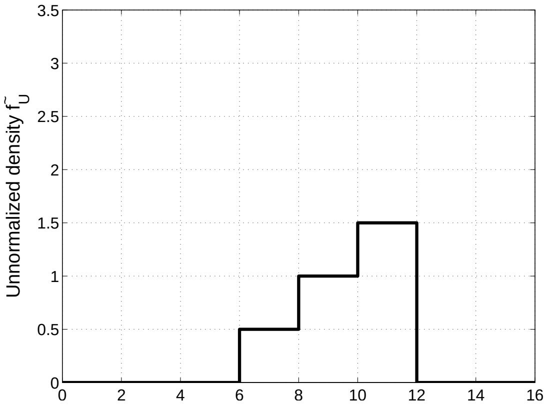
1.6 Let $U = 2X$ where X represents the original r.v in each of the four cases. Find S_U the set of values for which $f_U > 0$ and draw in the graphs below the (unnormalized) density of U for cases **b, d**. 2 p.

1.7 Compute the normalization constants for the two densities obtained in **1.6**; write their values above the respective graphs. 1 p.

(b)
$\mathbf{S}_U = [2, 8]$
Z=3



(d)
$\mathbf{S}_U = [6, 12]$
Z=6



Problem 2 6 points

Compute and make neatly labeled graphs of the density estimate for each of the (very small) data sets given below. If you choose to draw an unnormalized f write the value of the normalization constant next to the plot. Show all your work.

2.1 Find a normal density estimate for the data = $\{2, 1, 1\}$ by the Maximum Likelihood method. 3 p.
(A table with square roots of numbers 2–20 is available at the bottom of the page.) **2.2** Find a 3 p.

kernel density estimate for a square kernel with $h = 3$. Assume that the square kernel is

$$k(x) \;=\; \begin{cases} 1 & \text{if } x \in [-0.5, 0.5] \\ 0 & \text{otherwise} \end{cases}$$

data = $\{-1,\ 1,\ 2\}$

Solution

2.1

$$\begin{aligned} \mu &= \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+1+2}{3} = \frac{4}{3} \\ \sigma^2 &= \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1} = \frac{(1-4/3)^2 \times 2 + (2-4/3)^2}{2} = \frac{1}{3} \\ f(x) &= \sqrt{\frac{3}{2\pi}} e^{-\frac{(x-4/3)^2}{2/3}} \end{aligned}$$

plot should be here...

2.2

$$\begin{aligned} f(x) &= \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right) \\ &= \frac{1}{2\times 3}\left[k\left(\frac{x+1}{3}\right) + k\left(\frac{x-1}{3}\right) + k\left(\frac{x-2}{3}\right)\right] \\ &= \begin{cases} \frac{1}{9} & -2.5 \le x < -0.5 \\ \frac{2}{9} & -0.5 \le x < 2.5,\ x \neq 0.5 \\ \frac{1}{9} & 2.5 \le x < 3.5 \\ \frac{3}{9} & x = 0.5 \end{cases} \end{aligned}$$

..and here

Problem 3 3 points

Let A, B be events in a sample space S endowed with probability distribution P, and $0 < P(B) = q < 1$. Show that

$$A \perp B \quad \text{implies} \quad A \perp \overline{B}$$

In the above, $\overline{B} = S \setminus B$ and the symbol $\perp$ denotes probabilistic independence.

3.2 (this wasn't on the exam, but I left it here) X, Y are random variables, $S_X = \{0,\ 1\}$, $S_Y = \{0,\ 1,\ 2\}$, $P_Y(y) > 0$ for all $y \in S_Y$. Show by a counterexample that

$$P_X(x) = P_{X|Y}(x|0) \text{ for all } x \in S_X \quad \not\Longrightarrow \quad X \perp Y$$

Solution Let $p = P(A) = P(A|B),\ q = P(B),\ P(A|\bar{B}) = x$. We want to show that $x = p$. We have

$$P(A) \;=\; P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

$$\begin{aligned} p &= pq + x(1-q) \\ p(1-q) &= x(1-q) \implies p = x \end{aligned}$$

Another solution: **3.1** $A \perp B \rightarrow P(AB) = P(A) \times P(B)$

$$\begin{aligned} P(A\overline{B}) &= P(A) - P(AB) \\ &= P(A) - P(A) \times P(B) \\ &= P(A) \times (1 - P(B)) \\ &= P(A) \times P(\overline{B}) \quad \Rightarrow A \perp \overline{B}. \end{aligned}$$

3.2 X, Y are random variables, $S_X = \{0, 1\}$, $S_Y = \{0, 1, 2\}$, $P_Y(y) > 0$ for all $y \in S_Y$.One counter example will be

X	0	1	P_Y
$Y = 0$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
$Y = 1$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
$Y = 0$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{8}$
P_X	$\frac{1}{2}$	$\frac{1}{2}$	

Problem 4 8 points

Two continuous random variables X, Y taking values in $S_X = S_Y = [0, \infty)$ are described by the joint density

$$f_{XY}(x, y) = \lambda^2 e^{-\lambda(x+y)} \quad x, y \in [0, \infty),\ \lambda > 0$$

Let $U = X + Y$. Answer the following questions, showing all your reasoning.

Note: The univariate exponential distribution is $f(t) = \lambda e^{-\lambda t},\ t \in [0, \infty),\ \lambda \in (0, \infty)$. $E[t] = \frac{1}{\lambda}$, $Var(t) = \frac{1}{\lambda^2}$. $\int e^{-\lambda t} dt = -\frac{1}{\lambda} e^{-\lambda t}$.

4.1 Show that $X \perp Y$. 1 p.

4.2 Are X and U independent? 1 p.

4.3 What are the marginal densities f_X, f_Y ? 1 p.

4.4 What is the expectation $E[U]$? 1 p.

4.5 What is the variance $Var\, U$? 2 p.

4.6 What is the probability distribution of U? Either the density $f_U(u)$ or the CDF $F_U(u)$ are acceptable answers. 2 p.

Solution

4.1

$$f_{XY}(x, y) = \lambda^2 \times e^{-\lambda(x+y)}$$

$$
\begin{aligned}
&= \lambda e^{-\lambda x} \times \lambda e^{-\lambda y} \\
&= f_X(x) \times f_Y(y) \quad \Rightarrow X \perp Y
\end{aligned}
$$

4.2 X and U are not independent. Since if X is given, the value of U should be greater than X. The support of the distribution of U given X will depend on the given value of X. Obviously, X and U are not independent.

4.3
From **4.1**, we know that $X \perp Y$ and that the density $f_{XY} = f_X f_Y$. Hence, by separating the factors depending on x and y we obtain: $f_X(x) = \lambda e^{-\lambda x}$ on S_X and a similar result for f_Y.

One question you may ask is: how did I know that the correct normalization constants for f_X, f_Y are λ? In this case, we know it from the symmetry of f_{XY}. In the general case, when the function f_{XY} can be separated into two factors but is not symmetric w.r.t to x, y, i.e $f_{XY} = Cg(x)h(y)$, we can immediately see that that $g(x)$ and $h(y)$ have to be the unnormalized densities of X, Y. Hence, all that remains is to find the respective normalization constants.

Another, "longer", solution is here:

$$
\begin{aligned}
f_X(x) &= \int_0^\infty f_{XY}(x,y)dy = \lambda e^{-\lambda x} \\
f_Y(y) &= \int_0^\infty f_{XY}(x,y)dx = \lambda e^{-\lambda y}
\end{aligned}
$$

Therefore, the marginal distributions of X, Y are both exponential(λ). So $E(X) = E(Y) = \frac{1}{\lambda}$, and $Var(X) = Var(Y) = \frac{1}{\lambda^2}$.

4.4 $E[U] = E[X+Y] = E[X] + E[Y] = \frac{2}{\lambda}$.

4.5 $Var[U] = Var[X+Y] = Var[X] + Var[Y] + 2 \times Cov(X,Y) = \frac{2}{\lambda^2}$, since X, Y are independent, which implies $Cov(X,Y) = 0$.

4.6

$$
\begin{aligned}
F(u) &= P(U \le u) \\
&= \int_0^u \int_0^{u-y} f_{XY}(x,y)dxdy \\
&= \int_0^u \lambda e^{-\lambda y}(1 - e^{-\lambda(u-y)})dy \\
&= \int_0^u \lambda(e^{-\lambda y} - e^{-\lambda u})dy \\
&= 1 - e^{-\lambda u} - \lambda u e^{-\lambda u} \\
f_U(u) &= F'(u) = \lambda^2 u e^{-\lambda u}.
\end{aligned}
$$

Another solution: Since X, Y are independent, we can obtain the density of $X + Y = U$ by convolution.

$$
f_U(u) = \int_0^u f_X(x) f_Y(u-x)dx
$$

$$
\begin{aligned}
&= \int_0^u \lambda e^{-\lambda x} \lambda e^{-\lambda(u-x)} dx \\
&= \int_0^u \lambda^2 e^{-\lambda u} dx \\
&= \lambda^2 e^{-\lambda u} \int_0^u dx \\
&= \lambda^2 u e^{-\lambda u}
\end{aligned}
$$

Problem 5 7 points

Cryptographer Sofia is trying to open the entrance to the mysterious "treasure of Decebal" (a king who died exactly 1900 years ago). The door has a lock that can be opened by a combination from the following letters:

l_1	l_2	l_3
A	A	E
B	I	M
M	M	P
	R	R
		T

That is, any combination (l_1, l_2, l_3) can be entered in the lock, but only one of them, "the key"word, opens the door to the treasure.

Sofia has the following probabilistic knowledge about the letters forming the key.

- The distribution of the first letter l_1 is uniform.
- If the first letter is A, then the second letter l_2 is not a vowel, but it can be any of the consonants M,R with equal probability.
- If the first letter is not A, the second letter is a vowel and $P(l_2 = A | l_1 \neq A) = \frac{1}{4}$, $P(l_2 = I | l_1 \neq A) = \frac{3}{4}$.
- The distribution of the third letter l_3 given the previous two is

$$
P(\, l_3 \,|\, l_1 \, < \, l_2 \,) \; = \; \left\{ \begin{array}{ll} \text{T} & \text{w.p. } 0.6 \\ \text{E, M, P, R} & \text{w.p. } 0.1 \end{array} \right.
$$

$$
P(\, l_3 \,|\, l_1 \, > \, l_2 \,) \; = \; \left\{ \begin{array}{ll} \text{E} & \text{w.p. } 0.6 \\ \text{M, P, R, T} & \text{w.p. } 0.1 \end{array} \right.
$$

Here $l_1 < l_2$ means that l_1 is before l_2 in the alphabetical order (for instance A $<$ M). Answer showing all your reasoning.

5.1 What is the probability of the word "BIT"? 1 p.

5.2 What is the probability of the event "B*T", where * stands for any letter? 1 p.

5.3 What is the probability that the key word ends in T? 2 p.

5.4 Sofia meets the ghost of a symbologist, who guards the treasure. She convinces the ghost that she is there on a scientific mission and not to steal the treasure. The ghost whispers the key word in her ear, but Sofia only understands something like "..A..". But she is sure that the key word is one the following four MAT, MAP, BIT, ART. 3 p.

Denote by H the event *" Sofia understands "..A.." "*. The probability of H if the word contains an A is 0.9, and the probability of understanding "..A.." if the word doesn't contain A is 0.1. Given whether the word contains an A or not, H is independent of everything else.

Knowing this, can you help Sofia by finding which word has the highest probability to be the key to the treasure? (Note that you are NOT required to also calculate the value of this probability.)

Solution 5.1

$$\begin{aligned} P(BIT) &= P(l_3 = T | l_1 = B, l_2 = I) \times P(l_2 = I | l_1 = B) \times P(l_1 = B) && (1) \\ &= 0.6 \times \frac{3}{4} \times \frac{1}{3} = 0.15 && (2) \end{aligned}$$

5.2

$$\begin{aligned} P(B*T) &= P(BIT) + P(BAT) && (3) \\ &= 0.15 + P(l_3 = T | l_1 = B, l_2 = I) \times P(l_2 = I | l_1 = B) \times P(l_1 = B) && (4) \\ &= 0.15 + 0.1 \times \frac{1}{4} \times \frac{1}{3} = 0.15 + \frac{1}{120} = \frac{19}{120} && (5) \end{aligned}$$

5.3

$$\begin{aligned} P(**T) &= P(T | l_1 > l_2) \times P(l_1 > l_2) + P(T | l_1 < l_2) \times P(l_1 < l_2) && (6) \\ P(l_1 > l_2) &= P(BA) + P(MA) + P(MI) = \frac{1}{3}(1 + \frac{1}{4}) = \frac{5}{12} && (7) \\ P(l_1 < l_2) &= 1 - P(l_1 > l_2) = \frac{7}{12} && (8) \\ P(**T) &= 0.6 \times \frac{7}{12} + 0.1 \times \frac{5}{12} = \frac{47}{120} && (9) \end{aligned}$$

5.4 Denote by J the event { MAT, MAP, BIT, ART }. We need to find which of $P(MAT \mid H, J)$, $P(MAP \mid H, J)$, $P(BIT \mid H, J)$, $P(ART \mid H, J)$ is larger. Let's consider $P(MAT \mid H, J)$.

$$\begin{aligned} P(MAT \mid H, J) &= \frac{P(H|MAT, J) P(MAT|J)}{P(H|J)} && (10) \\ &= \frac{P(H|MAT) P(MAT)}{P(H|J) P(J)} && (11) \end{aligned}$$

because

$$P(MAT|J) = \frac{P(MAT)}{P(J)}$$

and because $P(H|MAT, J) = P(H|MAT)$, since H is independent of J given the word MAT.

Note now that the denominator of the (11) doesn't depend at all on the word MAT, and it's the same for all of the four words. So, to find which word has highest probability given H, J we only need to compute *the numerators*. Here they are:

$$\begin{aligned}
P(MAT \,|\, H, J) &\propto P(H|MAT)P(MAT) \\
&= P(H|\text{“contains A''})P(l_1 = M)P(l_2 = A|l_1 = M)P(T|l_1 > l_2) \\
&= 0.9 \times \frac{1}{3} \times \frac{1}{4} \times 0.1 \\
&= \frac{9}{1200}
\end{aligned}$$

$$\begin{aligned}
P(MAP \,|\, H, J) &\propto P(A|MAP)P(MAP) \\
&= 0.9 \times \frac{1}{3} \times \frac{1}{4} \times 0.1 \;=\; \frac{9}{1200}
\end{aligned}$$

$$\begin{aligned}
P(BIT \,|\, H, J) &\propto P(H|BIT)P(BIT) \\
&= P(H \,|\, \text{“does not contain A''})P(BIT) \\
&= 0.1 \times \frac{19}{120} \;=\; \frac{19}{1200}
\end{aligned}$$

$$\begin{aligned}
P(ART \,|\, H, J) &\propto P(H|ART)P(ART) \\
&= P(H \,|\, \text{“contains A''})P(l_1 = A)P(l_2 = R|l_1 = A)P(T|l_1 < l_2) \\
&= 0.9 \times \frac{1}{3} \times \frac{1}{2} \times 0.6 \\
&= \frac{9}{100}
\end{aligned}$$

Clearly the word with the largest probability is “ART”!