Probability for Statistics
Unseen Problem 3

1. In this question, we will see how to evaluate some definite integrals using only simple ideas from probability theory. Let $X_1, X_2, \ldots$ be a sequence of independent random variables, uniformly distributed on $(0, 1)$.

   Consider the first two variables, X_1 and X_2. Clearly, by symmetry, the probability that X_1 lies to the right of X_2 is $\frac{1}{2}$. This probability can also be calulated by conditioning on the value of X_1:

   $$\Pr(X_2 < X_1) = \int_0^1 \Pr(X_2 < x)\, dx = \int_0^1 x\, dx = \frac{1}{2}.$$

   (a) Use a similar argument to that given above to explain, without explicitly calculating the integral, why if n is a positive integer,

   $$\frac{1}{n+1} = \int_0^1 x^n\, dx.$$

   Both sides give the probability that X_1 is the largest amongst $X_1 \ldots X_{m+1}$. The integral conditions on $X_1 = x$ and integrates the probability density that $X_2, \ldots X_{m+1}$ are all smaller than x.

   (b) Extend your argument to calculate the beta integral, for positive integer m and n,

   $$\int_0^1 x^m (1-x)^n\, dx.$$

   This is the probability that X_1 is the $m+1$th value amongst $m+n+1$ uniform variables, so that there are m values smaller and n values larger. Hence

   $$\frac{m!n!}{(m+n+1)!} = \int_0^1 x^m (1-x)^n\, dx.$$

2. The data in Table 1 concern a long-term study into the benefits of screening for breast cancer. The subjects were 62,000 women in the USA who were members of a particular health insurance plan. The women were randomized to two groups of equal size. Women in the treatment group were encouraged to attend an annual screening, while those in the control group were offered their usual health care. Not all of the women in the treatment group attended the screening: 10,800 subjects refused. Less affluent subjects were more likely to refuse screening.

   (a) Is screening worthwhile? Explain with reference to the data in the table.

   Yes - looking at rate, we see 1.3 deaths per thousand in the treatment group versus 2.0 deaths per thousand in the control group.

   (b) Is breast cancer associated with income? If so, in what direction is the association?

   The refusers group, which has more subjects from low income backgrounds than the general population, has a lower mortality rate from breast cancer (1.5 per thousand) than the control group (2.0 per thousand). This suggests that breast cancer is more common in higher income individuals.

Table 1: Results from a study of screening for breast cancer in the USA. Rates are per thousand

		Cause of death			
		breast cancer		all other	
		number	rate	number	rate
Treatment group					
Examined	20,200	23	1.1	428	21
Refused	10,800	16	1.5	409	38
Total	31,000	39	1.3	837	27
Control group	31,000	63	2.0	879	28

(c) The death rate from all causes is roughly halved in the examined group compared with the group who refused treatment. Is screening responsible for this difference? If not, what explains the difference?

It is not reasonable to claim screening is responsible for the difference. The comparison is between groups that differ with respect to other factors relevant to mortality, such as income and education level.

Data taken from *Freedman, Pisani & Purves, Statistics*.