15-359: Probability and Computing

Assignment 6 Due: March 2, 2012

Problem 1: How we get them (10 pts.)

Let $X \sim \text{Uniform}(0,1)$. Prove that $-\ln X$ is an exponentially distributed variable. What is the parameter λ?

Problem 2: The rain in Spain (20 pts.)

The time until the next rainfall in a certain region has the memoryless property, assuming we can ignore the effects of changing seasons. Suppose that during the dry season, the time until the next rainfall (in days) in the central plain of the Iberian peninsula is $X \sim \text{Exp}(1/25)$.

A mild drought is defined as a period of at least 10 days without rainfall. We are interested in knowing more about the distribution of the lengths of such droughts.

A. Find $\mathrm{E}(X \mid X > 10)$ (the expected length of a mild drought) by integrating the conditional probability density function.

B. Find $\mathrm{E}(X \mid X > 10)$ by using the memorylessness property of the exponential distribution.

C. Find $\mathrm{E}(X^2 \mid X > 10)$ using any method you like. How does $\mathrm{Var}(X \mid X > 10)$ compare with $\mathrm{Var}(X)$?

Problem 3: Failure rate (15 pts.)

Let X be a continuous random variable with probability density function $f(t)$ and cumulative distribution function $F(t) = \mathrm{P}(X < t)$. We define the *failure rate* of X to be

$$r(t) := \frac{f(t)}{\overline{F}(t)} = \frac{f(t)}{1 - F(t)}.$$

Thus $r(t)\,dt$ represents the conditional probability that X is in the interval $(t, t+dt)$ given that $X > t$. If X represents the time to failure of some item, and t the current time, this is the probability that an item will fail in the next dt seconds, given that it has not failed so far.

A. Prove that the failure rate is a constant when X is distributed exponentially.

B. Prove that the exponential distribution is the only distribution with a constant failure rate.

Problem 4: (λ) (20 pts.)

One of the actual honest examples of exponential variables in real life is the decay of a radioactive isotope. Specifically, if a single unstable atom decays after T seconds, then $T \sim \text{Exp}(\lambda)$ for some λ.

A. Usually the unstability of an isotope is given not in terms of λ but rather the half-life $t_{1/2}$: the expected time for half of some amount of the isotope to decay. What is the relationship between λ and $t_{1/2}$?

B. Suppose that you have n atoms of an isotope, and the time it takes each one to decay is an independent variable $X_i \sim \text{Exp}(\lambda)$. Let $Y = \max\{X_1, \ldots, X_n\}$: the time it takes for all n atoms to decay. What is $\text{E}(Y)$? *(Hint: it may help to think about the case $n = 2$ first.)*

C. In most applications, the number of atoms, n, is mindbogglingly huge. Give an approximation of $\text{E}(Y)$ for large n, in units of half-lives.

Problem 5: Sparse selection (20 pts.)

Background

Suppose A is a set of cardinality n, $A \subseteq \mathcal{U}$ for some totally ordered set $\mathcal{U}$. Recall the clever randomized selection algorithm from class that computes $\mathsf{ord}(k, A)$ with probability $1 - \mathcal{O}(n^{-1/4})$. As in class, we refuse to write ceilings and floors.

1. Select a subset $B \subseteq A$ of cardinality $n^{3/4}$ at random (actually, sample with replacement).
2. Sort B.
3. Let $\kappa = k/n^{1/4}$, $\kappa^- = \max(\kappa - \sqrt{n}, 1)$, $\kappa^+ = \min(\kappa + \sqrt{n}, n^{3/4})$, $b^\pm = \mathsf{ord}(\kappa^\pm, B)$.
4. Compute $r^\pm = \mathsf{rk}(b^\pm, A)$ – note the A.
5. Let
$$A_0 = \begin{cases} \{\, x \in A \mid x \leq b^+ \,\} & \text{if } k < n^{1/4}, \\ \{\, x \in A \mid x \geq b^- \,\} & \text{if } k > n - n^{1/4}, \\ \{\, x \in A \mid b^- \leq x \leq b^+ \,\} & \text{otherwise.} \end{cases}$$
6. if $t \notin A_0$ or $|A_0| > 4n^{3/4}$ return to step 1.
7. Sort A_0 and return $\mathsf{ord}(k - r^- + 1, A_0)$.

In class we gave part of the proof that with probability $1 - \mathcal{O}(n^{-1/4})$ the algorithm terminates after the first round: we showed that $\mathsf{ord}(k, A) \in A_0$ with high probability. Here is the missing part of the proof: A_0 has cardinality bounded by $4n^{3/4}$, with high probability. Again, for simplicity, we only consider the case $\{\, x \in A \mid b^- \leq x \leq b^+ \,\}$; moreover, let's just say $k = n/2$.

Define the ranks $\lambda^- = \max(n/2 - 2n^{3/4}, 1)$, $\lambda^+ = \max(n/2 + n^{3/4}, n)$ and the corresponding points $a^\pm = \mathsf{ord}(\lambda^\pm, A)$.

Task

A. Argue that it suffices to check that $a^- \leq b^- < b^+ \leq a^+$ (in fact, this condition is stronger than the cardinality bound).

B. Using part (A), show that with probability $1 - \mathcal{O}(n^{-1/4})$ the cardinality bound holds for A_0.

Problem 6: Bayes of our lives (15 pts.)

Analogues of Bayes' theorem exist for continuous distributions as well. For instance, if X is a continuous distribution, we have

$$f_X(x \mid Y = y) = \frac{P(Y = y \mid X = x) f_X(X)}{P(Y = y)}.$$

Suppose the number of seasons in a television series is distributed as $N \sim \text{Geometric}(P)$: after each season, there is a fixed probability that the series is canceled. However, the parameter P depends on the popularity of the series, so we don't know what it is in general. We assume a uniform probability distribution for P: for a new series, $P \sim \text{Uniform}(0, 1)$.

A. A television series has been running for 47 seasons (and renewed for more). Write an expression for $f_P(p \mid N > 47)$ using Bayes' theorem.

B. What is $\mathrm{E}(P \mid N > 47)$? (This is the expected value of the show's probability of cancellation.)

C. **Extra credit** (15 pts.) What is the distribution of N, given that $N > 47$? What are its mean and median? *(You may want to use a CAS for these calculations.)*