Homework Set No. 8 – Probability Theory (235A), Fall 2011

Due: 11/22/11

1. (a) Let $X_1, X_2, \ldots$ be a sequence of independent r.v.'s that are uniformly distributed on $\{1, \ldots, n\}$. Define

$$T_n = \min\{k : X_k = X_m \text{ for some } m < k\}.$$

If the X_j's represent the birthdays of some sequence of people on a planet in which the calendar year has n days, then T_n represents the number of people in the list who have to declare their birthdays before two people are found to have the same birthday. Show that

$$\mathbf{P}(T_n > k) = \prod_{m=1}^{k-1} \left(1 - \frac{m}{n}\right), \qquad (k \geq 2),$$

and use this to prove that

$$\frac{T_n}{\sqrt{n}} \implies F_{\text{birthday}},$$

where F_{birthday} is the distribution function defined by

$$F_{\text{birthday}}(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-x^2/2} & x \geq 0 \end{cases}$$

(note: this is not the same as the normal distribution!)

(b) Take $n = 365$. Assuming that the approximation $F_{T_n/\sqrt{n}} \approx F_{\text{birthday}}$ is good for such a value of n, estimate what is the minimal number of students that have to be put into a classroom so that the probability that two of them have the same birthday exceeds 50%. (Ignore leap years, and assume for simplicity that birthdays are distributed uniformly throughout the year; in practice this is not entirely true.)

2. Consider the following two-step experiment: First, we choose a uniform random variable $U \sim U(0, 1)$. Then, conditioned on the event $U = u$, we perform a sequence of n coin tosses with bias u, i.e., we have a sequence $X_1, X_2, \ldots, X_n$ such that conditioned on the event $U = u$, the X_k's are independent and have distribution $\text{Binom}(1, u)$. (Note: without this conditioning, the X_k's are not independent!)

Let $S_n = \sum_{k=1}^n X_k$. Assume that we know that $S_n = k$, but don't know the value of U. What is our subjective estimate of the probability distribution of U given this information? Show that the conditional distribution of U given that $S_n = k$ is the beta distribution $\text{Beta}(k+1, n-k+1)$. In other words, show that

$$\mathbf{P}(U \le x \mid S_n = k) = \frac{1}{B(k, n-k)} \int_0^x u^k (1-u)^{n-k} \, du, \qquad (0 \le x \le 1).$$

Note: This problem has been whimsically suggested by Laplace in the 18th century as a way to estimate the probability that the sun will rise tomorrow, given the knowledge that it has risen in the last n days. (Of course, this assumes the unlikely theological scenario whereby at the dawn of history, a $U(0,1)$ random number U was drawn, and that subsequently, every day an independent experiment was performed with probability U of success, such that if the experiment is successful then the sun rises.)

Hint: Use the following density version of the total probability formula: If A is an event and X is a random variable with density f_X, then

$$\mathbf{P}(A) = \int_{\mathbb{R}} f_X(u) \mathbf{P}(A \mid X = u) \, du.$$

Note that we have not defined what it means to condition on a 0-probability event (this is a somewhat delicate subject that we will not discuss in this quarter) — but don't worry about it, it is possible to use the formula in computations anyway and get results.

3. Let $Z_1, Z_2, \ldots$ be a sequence of i.i.d. random variables with the standard normal $N(0,1)$ distribution. For each n, define the random vector

$$\mathbf{X}_n = (X_{n,1}, \ldots, X_{n,n}) = \frac{1}{\left(\sum_{i=1}^n Z_i^2\right)^{1/2}} (Z_1, \ldots, Z_n)$$

(a) The distribution of the random vector $\mathbf{X}_n$ is called the *uniform distribution on the* $(n-1)$*-dimensional sphere* $S^{n-1} = \{x \in \mathbb{R}^n : ||x|| = 1\}$. Explain why this makes intuitive sense, and if possible explain rigorously what conditions this distribution satisfies that justifies describing it by this name.

(b) Show that $\sqrt{n} X_{n,1} \implies N(0,1)$ as $n \to \infty$.

Hint. Use the law of large numbers.

(c) (Optional) For each $n \geq 1$, find the density function of the coordinate $X_{n,1}$.

Hint. Do it first for $n = 2$ and $n = 3$, and generalize using ideas from multivariate calculus. For $n = 3$, you should find that $X_{3,1} \sim U[-1, 1]$, a geometric fact which was known to Archimedes.

4. Compute the characteristic functions for the following distributions.

(a) **Poisson distribution:** $X \sim \text{Poisson}(\lambda)$.

(b) **Geometric distribution:** $X \sim \text{Geom}(p)$ (assume a geometric that starts at 1).

(c) **Uniform distribution:** $X \sim U[a, b]$, and in particular $X \sim [-1, 1]$ which is especially symmetric and useful in applications.

(d) **Exponential distribution:** $X \sim \text{Exp}(\lambda)$.

(e) **Symmetrized exponential:** A r.v. Z with density function $f_Z(x) = \frac{1}{2}e^{-|x|}$. Note that this is the distribution of the exponential distribution after being "symmetrized" in either of two ways: (i) We showed that if $X, Y \sim \text{Exp}(1)$ are independent then $X - Y$ has density $\frac{1}{2}e^{-|x|}$; (ii) alternatively, it is the distribution of an "exponential variable with random sign", namely $\varepsilon \cdot X$ where $X \sim \text{Exp}(1)$ and ε is a random sign (same as the coin flip distribution mentioned above) that is independent of X.