Introduction to Probability and Statistics - 18.05 Spring 2008

Test 2 - Solutions

1. Let n be the number of samples. Let X_i be the number of eggs in the i'th sample ($1 \leq i \leq n$). Then $E[X_i] = \mu$ where μ is unknown to us and $Var[X_i] \leq 3$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (i.e. the average of the samples). Then by linearity of expectation, $E[\bar{X}_n] = \mu$. Also, by independence, $Var[\bar{X}_n] = \frac{Var[X_i]}{n} \leq \frac{3}{n}$. We want that the probability that $|\bar{X}_n - \mu| \leq 2$ is at least $\frac{999}{1000}$, or that the complement of this event has probability at most $\frac{1}{1000}$. By Chebyshev inequality,

$$\Pr[|\bar{X}_n - \mu| > 2] \leq \frac{Var[\bar{X}_n]}{4} \leq \frac{3}{4n} \leq \frac{1}{1000}$$

   solving the last inequality for n, we get that $n \geq 750$. In other words, if the scientists take 750 samples they are guaranteed that their estimate will be within 2 of the real value with probability at least 999/1000.

2. (a) Linearity of expectation says that for random variables $X_1, \ldots, X_n$, and real numbers $a_0, a_1, \ldots, a_n$,

$$E[a_0 + \sum_{i=1}^{n} a_i X_i] = a_0 + \sum_{i=1}^{n} a_i E[X_i]$$

   It holds for any set of random variables (provided that they have finite expectations).

   (b) For every $1 \leq i \leq n$, define an indicator r.v. $Y_i = 1$ if the i'th egg was returned to its original nest and 0 otherwise. Then $\Pr[Y_i = 1] = E[Y_i] = \frac{1}{n}$. This is because each egg is equally likely to be placed in any one of the n nests. Let $Y = \sum_{i=1}^{n} Y_i$ be the number of eggs that are placed in their original nests. By linearity of expectation:

$$E[Y] = E[\sum_{i=1}^{n} Y_i] = \sum_{i=1}^{n} E[Y_i] = n \cdot \frac{1}{n} = 1$$

So in expectation, exactly one egg will be placed in its original nest.

3. (a) The Central Limit Theorem: let $X_1, X_2, \ldots$ be independent and identically distributed r.v.'s, each having expectation μ and variance σ^2. Then for every $-\infty < a < \infty$,

$$\lim_{n\to\infty} \Pr\left[\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le a\right] = \Phi(a)$$

   where $\Phi(a)$ is the value of the distribution function of the standard normal random variable at the point a.

   (b) Let X_i (for $1 \le i \le 10$) be the random variable that measures the time interval between the $(i-1)$'th eruption and the i'th eruption. We are interested in the r.v. $X = \sum_{i=1}^{10} X_i$, and want to estimate the probability that it is more than 1200. Since the X_i's are independent exponential r.v.'s with $\lambda = \frac{1}{100}$, we know that $E[X_i] = \frac{1}{\lambda} = 100$, and $Var[X_i] = \frac{1}{\lambda^2} = 10000$. Using the Central Limit Theorem (where Z is a standard normal r.v.):

$$\begin{aligned}\Pr[X > 1200] &= \Pr\left[\frac{X - 10\cdot 100}{\sqrt{10}\cdot 100} > \frac{1200 - 10\cdot 100}{\sqrt{10}\cdot 100}\right] \\ &\approx \Pr[Z > 0.63] = 1 - \Phi(0.63) = 0.2643\end{aligned}$$

4. (a) X, Y are not independent. Intuitively, because the distribution of Y by definition depends on the value of X. More formally, consider the events $X > 5$ and $Y > 5$. If they were independent it would hold that $\Pr[Y > 5 | X > 5] = \Pr[Y > 5]$. However, $\Pr[Y > 5 | X > 5] = 1$, since Y is always at least as large as X. On the other hand, as we will see later, $\Pr[Y > 5] < 1$ (even without calculating this probability, it is clear that it is not 1, since Y can take values smaller than 5 with positive probability).

   (b) The joint density function is $f_{X,Y}(x, y) = \frac{1}{10(10-x)}$ for $0 < x < 10$ and $x < y < 10$. Intuitively, this follows the definition of conditional probability:

$$\Pr[X = x \text{ and } Y = y] = \Pr[X = x] \cdot \Pr[Y = y | X = x]$$

by the fact that X is uniform on $(0, 10)$ and Y is uniform on $(X, 10)$, we get that the above is $\frac{1}{10(10-x)}$, when $0 < x < 10$ and $x < y < 10$. This is not a formal argument since X and Y are continuous (and hence we cannot talk about exact values x, y). However we can formalize this by looking at small intervals around these points.

(c) We divide the space into two events: $0 < X \leq 5$ and $X > 5$.

$$\Pr[Y > 5] = \int_0^5 \int_5^{10} \frac{1}{10(10-x)} dydx + \int_5^{10} \int_x^{10} \frac{1}{10(10-x)} dydx$$

The first integral is $\frac{\ln 2}{2}$. We can also compute the second integral, but instead we observe that when $X > 5$, with probability 1 $Y > 5$. So the second integral is just the probability that $X > 5$ which is $1/2$ (since X is uniform over $(0, 10)$). Hence we get that $\Pr[Y > 5] = \frac{\ln 2 + 1}{2}$.