Math 425, Section 6, Fall 1999: Introduction to Probability

Second Mid-Term: Solutions

1. Consider the following graph of the joint distribution of two random variables $\mathbf{X_1}, \mathbf{X_2}$.

		$\mathbf{X_2}$ 0	1	$p_{\mathbf{X_1}}(x_1)$
$\mathbf{X_1}$	0	$\frac{25}{102}$	$\frac{13}{51}$	
	1	$\frac{13}{51}$	$\frac{25}{102}$	
	$p_{\mathbf{X_2}}(x_2)$			1

(i) Fill out the table by finding the values of the distributions $p_{\mathbf{X_i}}(x_i), i = 1, 2$.
(ii) Are the random variables $\mathbf{X_1}$ and $\mathbf{X_2}$ independent?
(iii) What is a situation for which this distribution would be a good probability model? Why?

(i) All four probabilities are $\frac{1}{2}$ – add the rows or columns.

(ii) No: $P(X_1 = 1 \text{ and } X_2 = 1) = \frac{25}{102} \neq P(X_1 = 1)P(X_2 = 1) = \frac{1}{4}$.

(iii) Draw two cards from a standard deck of playing cards, without replacement. Let X_i be the indicator variable for drawing a red card on the i-th draw. Then the probabilities match exactly.

2. Five hundred independent rolls of a fair die will be made. What is the approximate probability that five will occur at least 100 times?

Use the normal approximation here (see your textbook, p. 212). You have a Binomial random variable, X, with mean $p = \frac{1}{6} \times n = 500 = 83.3$ and variance $np(1-p) = 500\frac{1}{6}\frac{5}{6} = 69.4$ and therefore standard deviation $\sigma = \sqrt{69.4} = 8.3$. We are lookiing for $P(X \geq 100) = P(\frac{X-83.3}{8.3} \geq \frac{100-83.3}{8.3}) = P(\frac{X-83.3}{8.3} \geq 2)$. The random variable $\frac{X-83.3}{8.3}$ is normalized to have mean 0 and variance 1. The normal approximation is to replace this by Z, the standard normal random variable. So we want to look up $P(Z \geq 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.02$. The last figures were gotten from the table for $\Phi(x)$ given out at the exam.

3. For a certain publisher, the number of misprints expected on any given page of a book is 0.1.

(a) What is the probability that you find an error on a given page of one of their books?

(b) If there is at least one misprint on this page, what is the probability that there are three on this same page?

Be sure to explain what probability distribution you are using to model this situation, and why.

Use a Poisson distribution with expectation $\lambda = 0.1$. This is because we view the event of finding an error to be a rare event, e.g., because there are many letters on a page and a small probability that any one will be erroneous.

(a) Let X be the random variable "the number of errors on the page". We assume, as above, $X \sim$ Poisson$[\lambda]$, with $\lambda = 0.1$. Then we want $P(X > 0) = 1 - P(X = 0) = 1 - e^{-0.1}$.

(b) Here we are interested in the conditional probability

$$P(X = 3 \mid X > 0) = \frac{P(X = 3 \text{ and } X > 0)}{P(X > 0)} = \frac{P(X = 3)}{P(X > 0)} = \frac{e^{-0.1}(0.1)^3/6}{1 - e^{-0.1}}.$$

4. (a) A pond contains 100 fish, of which 30 are carp. Subsequently, 20 fish are caught in a batch. Let $\mathbf{X}$ be the random variable, "the number of carp caught amongst these 20". What are the mean and the variance of $\mathbf{X}$? What assumptions are you making?

(b) A pond contains 100 fish, of which 30 are carp. Subsequently, a fish is caught, and it is recorded whether it is a carp or not. Then the fish is thrown back and another fish is caught, it is recorded whether it is a carp or not, and so on, until a fish has been caught, observed and returned 20 times. Let $\mathbf{Y}$ be the random variable, "the number of carp caught in those 20 times". What are the mean and the variance of $\mathbf{Y}$? What assumptions are you making?

(c) If you *did not know* how many carp there were in the pond, which method of sampling described in (a) and (b) above would probably be the more accurate as a way to estimate how many carp there were in the pond? Be sure to give a reason for your answer.

(a) Use the hypergeometric distribution, as on the examination question sheet. Here $N = 100, M = 30, n = 20$, and you find expectation $E(X) = 6$ and variance $Var(X) = 3.39$.

(b) Use the binomial distribution, with $n = 20, p = 0.3$, to get expectation 6 and variance 4.2.

(c) Use method (a) because the variance is less, and the sample will stray less from the expected value.

5. Let $\mathbf{X}$ be a random variable which is uniformly distributed over the interval from a to $b, a < b$, in the real line $\mathbb{R}$. Calculate the mean and variance of $\mathbf{X}$. Be sure to show all your work in this calculation.

The answers are given on the data sheet you were provided: the expectation is $\frac{a+b}{2}$ and the variance is $\frac{(b-a)^2}{12}$. The details of the calculation are on pages 201 and 202 of the textbook.