Denker **SPRING 2010**

416 Stochastic Modeling - Assignment 1

SOLUTIONS

Problem 1: (Problem 51, page 92) A coin, having probability p for landing heads, is flipped until head appears for the r-th time. Let N denote the number of flips required. Calculate $E[N]$.

Hint: Write N as a sum of geometric random variables.

Solution: Let X_j denote the waiting time for the j-th head to show up. We claim that X_j is a geometric random variable with parameter p. Indeed, if head appears for the $j-1$-th time, then flipping the coin is just like starting to flip a coin, so the probability of appearing head first in the n-th trial is the same as the probability of having a head the first time in the n-th trial after the $j-1$-th appearance of head.

So, $N = \sum_{i=1}^{r} X_i$ is the waiting time for the r-th head to show up and

$$EN = \sum_{i=1}^{r} EX_i = rEX_1 = \frac{r}{p}$$

is the expected time (see book page 68/69 for the expectation of a geometric random variable with parameter p).

Problem 2: (Problem 56, page 93) There are n types of coupons. Each newly obtained coupon is, independently, type i with probability p_i, $i = 1, ..., n$. Find the expected number and the variance of the number of distinct types obtained in a collection of k coupons.

Solution: Let ξ_l be a random variable which is 1 if the type l occurs in the collection of k coupons and be 0 otherwise. We are interested in the expectation and variance of $N = \sum_{i=1}^{n} \xi_i$.

Now

$$P(\xi_l = 0) = P(\text{type } l \text{ is not used in the collection}) = (1 - p_l)^k,$$

hence

$$EN = \sum_{l=1}^{n} E\xi_l = \sum_{l=1}^{n} (1 - (1 - p_l)^k).$$

Since for $l \neq l'$

$$\begin{aligned} 1 - P(\xi_l = 1, \xi_{l'} = 1) &= P(\{\xi_l = 0\} \cup \{\xi_{l'} = 0\}) \\ &= P(\xi_l = 0) + P(\xi_{l'} = 0) - P(\xi_l = 0, \xi_{l'} = 0) \\ &= (1 - p_l)^k + (1 - p_{l'})^k - (1 - p_l - p_{l'})^k \end{aligned}$$

we get

$$P(\xi_l = 1, \xi_{l'} = 1) = 1 - [(1 - p_l)^k + (1 - p_{l'})^k + (1 - p_l - p_{l'})^k].$$

The variance of N is

$$Var(N) = \sum_{i=1}^{n} Var(\xi_i) + 2 \sum_{j=2}^{n} \sum_{i=1}^{j-1} Cov(\xi_i, \xi_j),$$

so, using what has been shown above ($E(\xi_i) = E(\xi_i^2) = 1 - (1 - p_i)^k$ and $E(\xi_i \xi_j) = 1 - [(1 - p_l)^k + (1 - p_{l'})^k + (1 - p_l - p_{l'})^k]$) we arrive at

$$Var(N) = \sum_{i=1}^{n} (1 - p_i)^k (1 - (1 - p_i)^k) + 2 \sum_{1 \leq i, j \leq n} [(1 - p_i - p_j)^k - (1 - p_i)^k - (1 - p_j)^k].$$

Problem 3: (Problem 62, page 94) In deciding upon the appropriate premium to charge, insurance companies sometimes use the exponential principle, defined as follows. With X as the random amount that it will have to pay in claims, the premium charged by the insurance company is

$$P = \frac{1}{a} \ln(E[e^{aX}])$$

where a is some specific positive constant. Find P when X is an exponential random variable with parameter λ, and $a = \alpha\lambda$, where $0 < \alpha < 1$.

Solution:

$$\frac{1}{a} \ln Ee^{aX} = \frac{1}{\alpha\lambda} \ln \int_0^{\infty} \lambda e^{-\lambda(1-\alpha)x} dx = -\frac{1}{\alpha\lambda} \ln(1 - \alpha).$$

Problem 4: (Problem 74, page 95) Let $X_1, X_2, ...$ be a sequence of idependent, identically distributed continuous random variables. We say that a record occurs at time n if $X_n > \max(X_1, ..., X_{n-1})$. That is, X_n is a record if it is larger than each of $X_1, ..., X_{n-1}$. Show

1. $P\{\text{a record occurs at time } n\} = \frac{1}{n}$.
2. $E[\text{number of records by time } n] = \sum_{i=1}^{n} \frac{1}{i}$.

3. $Var(\text{number of records by time } n) = \sum_{i=1}^{n} \frac{i-1}{i^2}$.

4. Let $N = \min\{n : n > 1 \text{ and a record occurs at time } n\}$. Show $E[N] = \infty$.

Solution: Let ξ_i denote the random variable which is 1 if a record occurs at time i, and be 0 otherwise. It follows that

$$P(\xi_i = 1) = P(\max_{k<i} X_k < X_i).$$

Since the X_k are iid, $P(\max_{k<i} X_k < X_i) = P(\max_{k\le i, k\ne l} X_k < X_l)$ and

$$1 = \sum_{l=1}^{i} P(\max_{k\le i, k\ne l} X_k < X_l),$$

it follows that $E\xi_i = \frac{1}{i}$, hence $S = \sum_{i=1}^{n} \xi_i$, the number of records by time n satisfies

$$ES = \sum_{i=1}^{n} \frac{1}{i}.$$

Similarly

$$P(\xi_j = 1, \xi_j = 1) = \frac{1}{ij}$$

for $1 \le i < j \le n$, hence the variables are uncorrelated and

$$Var(N) = \sum_{i=1}^{n} \frac{1}{i} - \frac{1}{i^2} = \sum_{i=1}^{n} \frac{i-1}{i^2}.$$

The event that a record occurs at time n but not before at times $2, ..., n-1$ is the event

$$A_n = \{\max_{2\le k<n} X_k < X_1 < X_n\}.$$

Its probability is $P(A_n) = \frac{1}{n(n-1)}$ by similar reasonings as above, so

$$EN = \sum_{n=2}^{\infty} nP(A_n) = \sum_{n=2}^{\infty} \frac{1}{n-1} = \infty.$$

Problem 5: (Problem 76, page 96) Let X and Y be independent random variables with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2. Show that

$$Var(XY) = \sigma_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2.$$

Solution: Using independence and the definition of the variance

$$\begin{aligned} & Var(XY) = E(X^2Y^2) - (EXY)^2 = EX^2EY^2 - (EX)^2(EY)^2 \\ = \; & (EX^2 - (EX)^2)(EY^2 - (EY)^2) + (EX^2 - (EX)^2)(EY)^2 + (EX)^2(EY^2 - (EY)^2). \end{aligned}$$