36

MATH 505a **QUALIFYING EXAM** **February 1, 2006**

Last Name: ____________________ **First Name:** ____________________

ID#: ____________________ **Signature:** ____________________

Fall 2004, #2 → better explanation of problem

1. A *run* in a sequence of coin tosses is a maximal subsequence of consecutive tosses all having the same outcome; for example HHHTHHTTH has 5 runs. A biased coin, with $p = \mathbf{P}(\text{heads}) \in (0,1)$, is tossed n times. Write $q = 1 - p$. Let R_n be the number of runs in the first n tosses. Find exact formulas for

a) $\mu_n = \mathbf{E}R_n$ and

b) $\sigma_n^2 = \text{Var}(R_n)$.

HINT: pay careful attention to boundary effects-what happens at the start and end of the sequence of n tosses. Note that $\mu_1 = 1, \sigma_1^2 = 0$, and use this as a check on your answers. Note also that $\mathbf{P}(R_2 = 1) = p^2 + q^2$, $\mathbf{P}(R_2 = 2) = 2pq$, so $\mu(2) = 1 + 2pq$. ← μ_2

c) For the special case $p = 1/2$, the distribution of $R_n - 1$ is very well known distribution (e.g. Binomial, Poisson, Hypergeometric, Geometric, etc) NAME the distribution AND its parameter(s).

2. Assume the vector $\mathbf{X} = (X_1, \ldots, X_N)$ has a multivariate normal distribution $N(\mu, \mathbf{V})$, where μ is the vector of expected values and $\mathbf{V}$ is the covariance matrix. Let $c_1, \ldots, c_N$ be constants.

Find the distribution of $Y = \sum_{i=1}^{N} c_i X_i$.

3. Let $S_n = X_1 + \cdots + X_n, n \geqq 1$, be a random walk, where $EX_k = \mu$ and $Var(X_k) = \sigma^2$, $0 < \sigma^2 < \infty$.

+ (a) Find the covariance $Cov(S_n, S_m)$ and the correlation coefficient $\rho(S_n, S_m)$ of S_n and S_m, $m \neq n$.

(b) Assume $n > m$. Find $\lim_{n\to\infty} Cov(S_n, S_m)$ and $\lim_{n\to\infty} \rho(S_n, S_m)$. Does $\lim_{n\to\infty} Cov(S_n, S_m)$ depend on the distribution of the increments? Does $\lim_{n\to\infty} \rho(S_n, S_m)$ depend on the distribution of the increments?

Spring 2006, Applied Probability.

1. Define A_i as the event that i^{th} and $(i+1)^{st}$ flip gives different outcomes (i.e. either HT or TH). So, $R_n = 1 + \sum_{i=1}^{n-1} 1_{A_i}$. Then,

a) $E[R_n] = 1 + \sum_{i=1}^{n-1} P(A_i) = 1 + \sum_{i=1}^{n-1} (pq + qp) = 1 + 2(n-1)pq$

b) $Var(R_n) = \sum_{i=1}^{n-1} var(1_{A_i}) + \sum_{1 \le i < j \le n-1} Cov(1_{A_i}, 1_{A_j})$

$\rightarrow$ $Var(1_{A_i}) = P(A_i) - P(A_i)^2 = 2pq(1-2pq)$

$\rightarrow$ If $j = i+1$: $Cov(1_{A_i}, 1_{A_{i+1}}) = P(A_i \cap A_{i+1}) - P(A_i)P(A_{i+1}) = (pqp + qpq) - 4p^2q^2$

$= pq(p+q-4pq)$

$\rightarrow$ If $j \geq i+2$: $Cov(1_{A_i}, 1_{A_j}) = P(A_i \cap A_j) - P(A_i)P(A_j) = 0$ because of independence of trials.

So, $Var(R_n) = 2(n-1)pq(1-2pq) + (n-2)pq(p+q-4pq)$

c) Suppose $p = q = 1/2$. Consider $P_n := R_n - 1 = \sum_{i=1}^{n-1} 1_{A_i}$. We know that when $|j-i| \geq 2$, 1_{A_i} and 1_{A_j} are independent. When $|j-i| = 1$, wlog when $j = i+1$, 1_{A_i} and 1_{A_j} are two random variables both taking only two values and $Cov(1_{A_i}, 1_{A_j}) = \frac{1}{4}(1 - 4 \cdot \frac{1}{4}) = 0$. So, they are actually independent. That is, P_n is summation of independent Bernoulli r.v. which means P_n has Binomial distribution.

$E[P_n] = 1 + 2(n-1) * \frac{1}{4} - 1 = \frac{n-1}{2} \quad (= (n-1)p)$

$Var(P_n) = 2(n-1) * \frac{1}{4}(1 - \frac{1}{2}) + (n-2) * \frac{1}{4}(1 - 4 * \frac{1}{4}) = \frac{n-1}{4} \quad (= (n-1)pq)$

Thus $P_n \sim Binomial(n-1, 1/2)$

2. Since X has a multivariate normal distribution, we know that,

$X_i = \mu_i + \sum_{j=1}^{N} a_{ij} Z_j$ for $i = 1, \ldots, N$ where $Z_1, \ldots, Z_N$ are iid $N(0,1)$. So,

$Y = \sum_{i=1}^{N} c_i \mu_i + \sum_{i=1}^{N} \sum_{j=1}^{N} c_i a_{ij} Z_j$ which means Y has normal distribution.

$$E[Y] = \sum_{i=1}^{N} c_i \mu_i = c^T \mu$$

$$Var(Y) = \sum_{i=1}^{N} c_i^2 V_{ii} + \sum_{1 \le i < j \le N} c_i c_j V_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j V_{ij} = c^T V c$$

Thus, $Y \sim Normal(c^T \mu, c^T V c)$.

2. a) Suppose wlog that $n > m$.

$$Cov(S_n, S_m) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, X_j) = \sum_{j=1}^{m} Var(X_j) = m\sigma^2$$

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) = n\sigma^2$$ and similarly, $Var(S_m) = m\sigma^2$

$$\rho(S_n, S_m) = \frac{Cov(S_n, S_m)}{\sqrt{Var(S_n)}\sqrt{Var(S_m)}} = \frac{m\sigma^2}{\sqrt{n}\sigma\sqrt{m}\sigma} = \frac{\sqrt{m}}{\sqrt{n}}$$