ACM116 - Fall 2007 - Final Exam Solutions

TA: Mulin Cheng, Roger Donaldson

Handed out: 10 Dec 2007, Due: 14 Dec 2007

1. (15 Points) You bet only when there is at least one even dice, so

$$\mathbb{P}(\text{I win}) = \mathbb{P}(\text{two evens}|\text{at least one even}) = \frac{\mathbb{P}(\text{two evens})}{\mathbb{P}(\text{at least one even})} = \frac{\frac{1}{2}\times\frac{1}{2}}{1-\frac{1}{2}\times\frac{1}{2}} = \frac{1}{3} \tag{1}$$

   and

$$\mathbb{P}(\text{I lose}) = 1 - \mathbb{P}(\text{I win}) = \frac{2}{3} \tag{2}$$

   Clearly, the game is not fair due to no betting when there is no even dice.

2. (15 Points) Strictly speaking, we should model the number of incidents, N, as a binomial random variable, but since $N \gg 1$ and $p \ll 1$, we can model the number of incidents by a Poisson variable of mean $\lambda = 81$. Furthermore, since λ is large, we can approximate the Poisson variable by a Gaussian of mean and variance λ. As a Gaussian variable, $N = 117$ is $z = (117 - 81)/9 = 4$ standard deviations away from the mean, and we conclude that this number of incidents is exceptionally high. Note that while it is possible to compute $\mathbb{P}(N \geq 117)$ from the Poisson approximation, using a Gaussian approximation is a bit more clever since it doesn't involve evaluating a nasty sum of Poisson terms, and also shows a more complete understanding of the distributions involved.

3. (15 Points) There are (at least) two ways to do this problem. A standard way is to assume that Aspirin has no effect, and to model the placebo data as a sample of a normal random variable. In such case the occurrence of a heart attack in each untreated individual is a Bernoulli variable with mean $p_0 = 209/11000 = 0.019$ and variance $p_0(1-p_0) = 0.0186$. Applying the central limit theorem, the fraction of heart attacks in the placebo group is a Gaussian variable with mean $p_0 = 0.019$ and variance $p_0(1-p_0)/n = 1.69 \times 10^{-6}$. Under the assumption that the treated group is sampled from the same distribution as the placebo group, the probability of finding a fraction of $104/11000 = 9.45 \times 10^{-3}$ or fewer heart attacks is

$$\Phi\left(\frac{9.45\times 10^{-3} - 1.9\times 10^{-2}}{\sqrt{1.69\times 10^{-6}}}\right) = \Phi(-7.33) < 10^{-11}, \tag{3}$$

   so we can conclude that it is unlikely that Aspirin has no effect on the frequency of heart attacks.

   A second way to approach the problem is to note that the numbers of heart attacks are far smaller than the sample sizes, and may be modeled by Poisson variables. The placebo group is a Poisson variable of mean 209. Proceeding as in problem 2, approximating the Poisson variable by a Gaussian variable, we find that the number of heart attacks in the treated group is $(209-104)/\sqrt{209} = 7.29$ standard deviations less than the placebo group. Again, we state that if the placebo group and treated group have the same distribution of heart attacks, for the treated group to have so few heart attacks is an unlikely event. We may conclude that it is unlikely that treatment by Aspirin has no effect on the frequency of heart attacks.

4. (20 Points) The expected value of Gaussian vector X is

$$\mu = \begin{pmatrix} 7.5\% \\ 10\% \\ 20\% \end{pmatrix} \tag{4}$$

and the covariance matrix is

$$\Sigma = \begin{pmatrix} 0.0049 & 0.0059 & -0.0063 \\ 0.0059 & 0.0144 & -0.0065 \\ -0.0063 & -0.0065 & 0.0324 \end{pmatrix} \tag{5}$$

Write vector $f = (f_A, f_B, f_C)^T$ as the fractions of investment on stock A, B, C, respectively, and then $1 - 1^T f$ is the fraction of investment on asset. Vector f uniquely determines the portfolio. Clearly, the return rate R of the portfolio is $f^T X + (1 - 1^T f) \times 5\%$.

(a) By the definition of Gaussian vector, the return rate R is a Gaussian random variable with mean $f^T \mu + (1 - 1^T f) \times 5\%$ and variance $f^T \Sigma f$. Here, the portfolio $f = (0.2, 0.2, 0.4)^T$. We find the expected return 12.5% and standard deviation 6.62%

(b) This problem can be formulated as an optimization problem as follows

$$\begin{aligned} \text{minimize} \quad & f^T \Sigma f && \text{(6a)} \\ \text{subject to} \quad & f_A, f_B, f_C \geq 0 && \text{(6b)} \\ & f_A + f_B + f_C \leq 1 && \text{(6c)} \\ & \mu^T f + (1 - 1^T f) 5\% \leq 12.5\% && \text{(6d)} \end{aligned}$$

The first two inequalities (6b) and (6c) come from the fact that f_A, f_B and f_C are fractions of investment on stocks. The last inequality comes from the requirement that the expected return rate is no less than that of the first portfolio. These three inequalities define a set in $\mathbb{R}^3$. Since we only want to find a better portfolio, we can consider minimizing the variance over a subset of the set defined by these three inequalities. Now suppose the last two inequalities are actually equalities. We can solve f_A and f_B from these two linear equations.

$$\begin{aligned} f_A &= 4 f_C - 1 && \text{(7a)} \\ f_B &= -5 f_C + 2 && \text{(7b)} \end{aligned}$$

Substitute these into the variance and get

$$\text{Var} R = f^T \Sigma f = 0.25 f_C^2 - 0.18764 f_C + 0.03898 \tag{8}$$

This is a quadratic function of f_C which takes minimum 0.0037712 at $f_C = 0.3753$. Therefore, the portfolio $f_A = 50.11\%$, $f_B = 12.35\%$ and $f_C = 37.53\%$ has expected return rate 12.5% with smaller standard deviation 6.14%. *(Note: This portfolio is also the best one.)*

(c) We know if $X \sim \mathcal{N}(\mu, \Sigma)$ is partitioned as

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) \tag{9}$$

Then the conditional distribution

$$X_2 | X_1 = x_1 \sim \mathcal{N}\left(\mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1), \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right) \tag{10}$$

Therefore, the distribution of the return rate of stock C conditioning on $R_A = -5\%$ and $R_B = 30\%$ is normal with expectation 41.22%

5. (15 Points) Under some conditions, we can use the following formula to find $\mathbb{P} N = n | T = t$

$$\mathbb{P}(N = n | T = t) = \lim_{\delta t \to 0} \frac{\mathbb{P}(N = n, T \in (t, t + \delta t))}{\mathbb{P}(T \in (t, t + \delta t))} \tag{11}$$

Now suppose δt is very small and write $M(t)$ as the number of shocks by time t. Obviously,

$$\mathbb{P}(T \in (t, t + \delta t)) = \sum_{m=0}^{\infty} \mathbb{P}(M(t) = m, T \in (t, t + \delta t)) \tag{12}$$

The event $\{M(t) = m, T \in (t, t + \delta t)\}$ means there are m shocks by time t which do not cause the system to fail and one shock in $(t, t+\delta t)$ which causes the system to fail. By the independent increment of Poisson process,

$$\mathbb{P}(T \in (t, t + \delta t)) = \sum_{m=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^m}{m!} (1 - p)^m \cdot (\lambda \delta t)p = \lambda p \delta t e^{-\lambda p t} \tag{13}$$

By similar argument, we have

$$\mathbb{P}(N = n, T \in (t, t + \delta t)) = e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} (1 - p)^{n-1} \cdot (\lambda \delta t)p \tag{14}$$

We then have the desired result

$$\mathbb{P}(N = n | T = t) = e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^{n-1}}{(n-1)!}. \tag{15}$$

This is valid only for $n > 0$: for $n \leq 0$, $\mathbb{P}(N = n|T) = 0$.

6. (20 Points)

   (a) Since $m \ll 200$, we may approximate the number of seeds that mature to new plants, M, by a Poisson variable. Furthermore, on average, only half of these seeds inherit the mutant gene, so the Poisson variable has mean $\lambda = m/2$. Therefore

   $$\mathbb{P}(M = k) = \frac{e^{-m/2}(m/2)^k}{k!}. \tag{16}$$

   (b) Let M_i be the number of mature mutants that are offspring of parent i, and let Z_n be the number of mutant parents in generation n. Then, assuming there are relatively few mutants (so it is unlikely that a seed has two mutant parents),

   $$Z_{n+1} = \sum_{i=1}^{Z_n} M_i. \tag{17}$$

   Note that $Z_1 = M_i$, and the $M_i \sim \textit{Poisson}(m/2)$ are iid. Compute the probability generating function for Z_{n+1} by conditioning on Z_n:

   $$\begin{aligned}
   \phi_{Z_{n+1}}(t) &= \mathbb{E}(t^{Z_{n+1}}) && (18)\\
   &= \mathbb{E}(\mathbb{E}(t^{Z_{n+1}}|Z_n)) && (19)\\
   &= \mathbb{E}((\mathbb{E}(t^{M_i}))^{Z_n}) \quad \text{(independence of the } M_i\text{)} && (20)\\
   &= \mathbb{E}((\phi_{Z_1}(t))^{Z_n}) && (21)\\
   &= \phi_{Z_n}(\phi(t)). && (22)
   \end{aligned}$$

   Hence, we have a recursive relationship for ϕ_{Z_n}, which is n compositions of $\phi(t) = \phi_{Z_1}(t)$.
   We want to compute $\mathbb{P}(Z_n = 0) = \phi_{Z_n}(0)$. Starting from $t = 0$, we compute the nth composition of ϕ's by an iteration between the curves $u = \phi(t)$ and $u = t$. See Figure 1. From the figure, we see that as long as $u = \phi(t)$ does not intersect $u = t$ for $t \in (0, 1)$, $\phi_{Z_n}(0) \to 1$ as $n \to \infty$ and the mutants will die out with probability 1. The marginal case occurs where $\phi'(1) = 1$. For a Poisson variable, $\phi(t) = \exp(\lambda(t - 1))$, so provided $\lambda > 1$, the mutant gene will survive. Hence, we need $m > 2$ for the gene to survive. When $m > 2$, the probability of survival is

   $$\begin{aligned}
   s^* &= 1 - \text{root of} \left(e^{\frac{m}{2}(t-1)} = t\right) && (23)\\
   &= \text{root of} \left(e^{-\frac{m}{2}s} = 1 - s\right), && (24)
   \end{aligned}$$

   where we choose the root that lies strictly on $0 < s < 1$. Given these two results,

   $$\mathbb{P}(\text{survival}) = \mathbb{P}(Z_n > 0, \forall n) = \begin{cases} 0, & m \leq 2, \\ s^*, & m > 2 \end{cases} \tag{25}$$

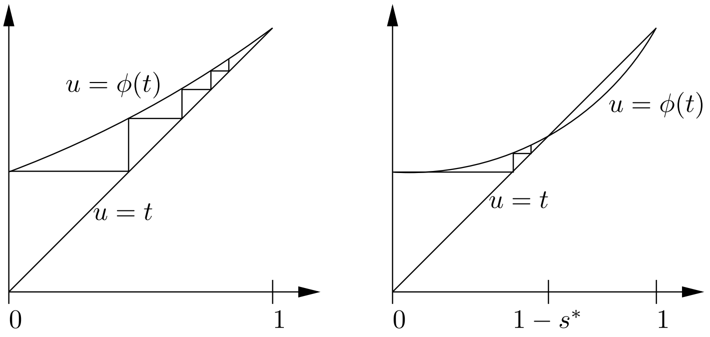

Figure 1: Determination of $\mathbb{P}(Z_n = 0)$ for $n \to \infty$. The iteration on the left has $m < 2$, whereupon $\mathbb{P}(Z_n = 0) \to 1$, and on the right, where $m > 2$, $\mathbb{P}(Z_n = 0) \to s < 1$.