Solutions to Some More Problems

Math/Stat 394

1 Quizzes

1.1

In the following list, mark the statements that are *possible*, leaving those that are impossible blank.

⊗ $EX = 3, EX^2 = 9$ It's a limit case, but it results in $Var[X] = 9 - 3^2 = 0$, which corresponds to a "degenerate" random variable, but is still legal.

◯ $EX^3 = -2$, EX does not exist - this cannot be: in general, if $E[X^n]$ is well defined, then all moments $E\left[X^k\right], k < n$ are also well defined.

⊗ $EX = -2, EX^2 = 5$ This presents no problem at all ($Var[X] = 5 - (-2)^2 = 1$)

◯ $EX = 0, Var[X] = -1$ Obviously, we cannot have a negative variance

1.2

In the following list, mark the statements that are *always* true.

◯ $E[XY] = E[X]E[Y]$ This is true only if X and Y are *uncorrelated* (for example, if they are independent)

⊗ $E[X+Y] = E[X] + E[Y]$ This is linearity of the expectation

◯ $E\left[\frac{X}{Y}\right] = \frac{E[X]}{E[Y]}$

◯ $E\left[\sqrt{X}\right] = \sqrt{E[X]}$

The last two are false, unless X is a constant.

1.3

Let X be a continuous RV, with a density $f_X(x)$ that is equal to zero for $x < 0$, and positive for $x \geq 0$. Let F_X be the corresponding cdf. In the following list mark the statements that are *certainly true*.

⊗ $F_X(0) = 0$ From what we know, $\int_{-\infty}^{x} f_X(t)\, dt = 0$ for $x \leq 0$

⊗ $F_X(-1) = 0$ See above

◯ $F_X(1) = 0$ On the other hand, $f_X(x) > 0$ for $0 \leq x \leq 1$, so the cdf, as its antiderivative, cannot be zero.

◯ $\lim_{x\to\infty} F_X(x) = 0$ This is never true.

◯ $\lim_{x\to\infty} F_X(x) = \frac{1}{2}$ Same here.

⊗ $\lim_{x\to\infty} F_X(x) = 1$ This is always true, regardless of any detail about the density.

2 Problems

2.1

Traffic flows into a highway from two separate feeders. A simple model is the following: we look at small time intervals of length h (for example 10 seconds), and the probability of one car coming in from feeder i ($i = 1, 2$) in each time interval is $\lambda_i \cdot h$, where $\lambda_i > 0$ is a constant. Assume each feed cannot produce more than one car in each time interval. We consider the number of cars entering the highway over a fixed time interval T (e.g., $T = 1$ hour).

1. What is the probability mass function for the number of cars entering from feeder i ($P[N_i = k]$, where N_i is the number of cars entering from feeder i)?
2. What is the probability mass function for the total number of cars entering the highway, $N = N_1 + N_2$
3. In the limit $h \to 0$, what do these three pmf's become?

Solutions

1. Since this is a binomial model

$$P[N_i = k] = \binom{1 \cdot \frac{1}{h}}{k} (\lambda_i \cdot h)^k (1 - \lambda_i \cdot h)^{\frac{1}{h} - k}$$

(we are assuming that $\frac{1}{h}$ is an integer - the number of time intervals of length h in the time between 0 and $T = 1$)

2. Adding up two independent binomially distributed random variable, does not, in general result in a binomial random variable. One way to see this is to consider the number of trials to have a success: being independent, the probability that the neither happens (hence, the distribution of the first combined success) is going to be determined (let's write $p_i = \lambda_i \cdot h$ for short) by the geometric distributions of the first successes (call the time interval of first arrival T_i)

$$P\left[\min\left(T_1, T_2\right) > n\right] = P\left[T_1 > n, T_2 > n\right] = P\left[T_1 > n\right] P\left[T_2 > n\right] =$$

$$= \left(1 - p_1\right)^n \left(1 - p_2\right)^n = \left(1 - p_1 - p_2 + p_1 p_2\right)^n$$

which corresponds to a geometric distribution with parameter $p_1 + p_2 - p_1 p_2$. On the other hand, the probability of both feeds producing one car at the same time is, by independence, $p_1 p_2 = \lambda_1 \lambda_2 h^2$. The probability of one car, from either feed, is them $p_1 + p_2 - 2p_1 p_2$. Since all trials are independent, once we have an arrival, we start again from scratch, and by successive arrival times, we have that the sum over a fixed amount of time will be multinomial. If we neglect the term in h^2, this reduced to a binomial distribution, with parameter $\left(\frac{1}{h}, p_1 + p_2\right)$.

3. In the limit $h \to 0$, the binomial distributions will turn into Poisson. By definition of the Poisson distribution,

$$P\left[N_i = k\right] = \frac{\left(\lambda_i \cdot 1\right)^k}{k!} e^{-\lambda_i \cdot 1} = \frac{\lambda_i^k}{k!} e^{-\lambda_i}$$

As for the minimum, it will converge to a binomial with parameter $\lambda_1 + \lambda_2$, since the term $p_1 p_2 = \lambda_1 \lambda_2 h^2$ will vanish in the limit as $h \to 0$. With a similar argument as the one in point 2, adapted to the Poisson case, which corresponds to a exponential distribution for the time of first arrival, one can show directly that the sum of two independent Poisson variables is distributed as Poisson, with a parameter equal to the sum of the parameters.

2.2

A simplified (but actually used in practice) stock market model is as follows. Starting from an initial price of $X_0 > 0$, a stock changes its price at every trade by moving, at the nth trade, from X_{n-1} up to $X_n = uX_{n-1}$ $(u > 1)$. or down to $X_n = dX_{n-1}$ $(d < 1)$. We assume that the move will be up with probability p, and down with probability $1 - p$. For simplicity we assume $u = 1 + c, d = 1 - c, 0 < c < 1$.

1. Consider X_m, the price after m trades. Its distribution is not one of the standard ones listed in the book, but has a fairly simple connection to a very standard distribution. Describe the distribution of X_m by a formula or in some other *precise* way.

2. Assume now that m is very large (mathematically, we are thinking of $m \to \infty$), and c very small (that is, $c \to 0$). How can we approximate the distribution of X_m? Note that, to make this into a precise mathematical limit, we need to assume that the rates for m and c are strictly related.

Hint: The easiest way to handle these questions is to look at the sequence $Y_0, Y_1, \ldots, Y_m$, with $Y_k = \log(X_k)$, that is $X_k = e^{Y_k}$

Solutions

1. Since $\log(X_k)$ takes two values, $\log u > 0$, and $\log d < 0$, this is a variation on the Bernoulli distribution. Specifically, the up movements follow a Bernoulli distribution with parameter p, and the down movements, one with parameter $1-p$, with the result being the difference between the two. Over a total of m trades, the result will be $k \cdot \log u + (m-k) \cdot \log d$, with probability $\begin{pmatrix} m \\ k \end{pmatrix} p^k (1-p)^{m-k}$

2. Since we are not touching the probability p, this calls for the Central Limit Theorem. For this to work, we need the number of steps, m, and the amount of each step, c, to be connected. First note that the expected logarithm of the price after m steps is $\log X_0 + m(p \log u + (1-p) \log d)$. In the limit of small c, assuming $u = 1+c, d = 1-c$, we have $(\log(1+x) \approx x)$
$$E \log X_m \approx \log X_0 + mc(2p-1)$$
On the other hand, the variance will be
$$Var[\log X_m] = m(\log u - \log d)^2 p(1-p) \approx 4mc^2 p(1-p)$$
The standard way is to assume that, instead of a fixed size c, we have a variable size keeping $c^2 m$ (approximately) constant (the same will then hold for $\sqrt{m}c$). Let's write $\sigma^2 = 4mc^2 p(1-p)$ and $\sqrt{m} \cdot \sqrt{m}c(2p-1) = \sqrt{m} \cdot \mu$ for short. The CLT governs the distance between $\log X_m$ and its expected value:
$$\lim_{m \to \infty} P\left[\frac{\log X_m - \sqrt{m}\mu - \log X_0}{\sigma} \le x\right] = \Phi(x)$$
that is, applying the binomial Central Limit Theorem, the result will be approximately Gaussian. Going back to the product of the X_k, that will be approximately *lognormal*: the density of $\log X_m - \log X_0$ being approximately normal with mean $\sqrt{m}\mu$ and variance σ^2, the ratio final price/initial price will have an approximate distribution
$$P\left[\frac{X_m}{X_0} \le x\right] = P[\log X_m - \log X_0 \le \log x] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\log x} \exp\left\{\frac{(u - \sqrt{m}\mu)^2}{2\sigma^2}\right\} du$$

The assumption of independence in the succeeding price steps is crucial, so this model is limited in this respect to situations where "herd effects" are absent. It is the basis of the famed Black-Scholes Formula, for the pricing of "European" options. Since the log-normality assumption leads to an explicit formula (involving Φ, of course), brokers used to have calculators at hand, with the formula pre-programmed.