# SOME COMPLEMENTARY MATERIAL TO THE COURSE "FINANCIAL RISK" MVE220/MAS400GU 

ALEXANDER HERBERTSSON

## Contents

| 1. | Introduction |  |
| :--- | :--- | :--- |
| 2. | The central limit theorem | 1 |
| 3. | The normal approximation of the binomial distribution | 1 |
| 4. | The normal approximation of the mixed binomial distribution | 2 |
| 5. | The law of large numbers approximation of a mixed binomial distribution |  |
| 6. | Finding the mixed binomial distribution bv using Monte Carlo simulations | 4 |
| 6.1. | Simulating $X_{1} \ldots, X_{m}$ in the mixed Merton binomial model | 6 |
|  |  | 8 |

## 1. Introduction

This short note is written for the students in the course "Financial Risk" MVE220 / MAS400GU at Chalmers university of technology and University of Gothenburg and treats some complementary material not covered in the lecture slides. The note considers the normal approximation of the mixed binomial model using the central limit theorem as well as Monte Carlo simulation of mixed binomial models. The material is not necessary restricted to credit risk and can be applied to any situation that requires mixed binomial models. Students in the course should read this paper carefully before solving task 2.2,2.3 and 2.4 in the credit risk project in the course "Financial Risk" MVE220 / MAS400GU.

## 2. The central limit theorem

Let $X_{1}, X_{2}, \ldots, X_{m}$ be independent and equally distributed random variables where $\mathbb{E}\left[X_{i}\right]=\mu$ and $\operatorname{Var}(X)=\sigma^{2}<\infty$ and define the so called sample mean $\bar{X}_{m}$, given by

$$
\begin{equation*}
\bar{X}_{m}=\frac{1}{m} \sum_{i=1}^{n} X_{i} . \tag{2,1}
\end{equation*}
$$

and note that $\mathbb{E}\left[\bar{X}_{m}\right]=\mu$ for any integer $m$. By applying Chebyshev's inequality to the random variable $\bar{X}_{m}$ we get

$$
\mathbb{P}\left[\left|\bar{X}_{m}-\mu\right| \geq \varepsilon\right] \leq \frac{\operatorname{Var}\left(\bar{X}_{m}\right)}{\varepsilon^{2}}=\frac{\operatorname{Var}\left(\frac{1}{m} \sum_{i=1}^{n} X_{i}\right)}{\varepsilon^{2}}=\frac{m \sigma^{2}}{m^{2} \varepsilon^{2}}=\frac{\sigma^{2}}{m \varepsilon^{2}}
$$

Date: November 26, 2012.
and we conclude that $\mathbb{P}\left[\left|\bar{X}_{m}-\mu\right| \geq \varepsilon\right] \rightarrow 0$ as $m \rightarrow \infty$. Note that this holds for any $\varepsilon>0$. This result is simple the law of large numbers which says that $\bar{X}_{m}$ converges (in probability) towards the mean $\mathbb{E}\left[X_{i}\right]=\mu$. Assume now that we want to find a sharper estimate of probability of the convergence rate of $\bar{X}_{m}$ towards the constant $\mu$ as $m \rightarrow \infty$. This can be done by using the so called central limit theorem (CLT) studied in your first statistic course.

Theorem 2.1. The central limit theorem (CLT) Let $X_{1}, X_{2}, \ldots, X_{m}$ be independent and equally distributed random variables where $\mathbb{E}\left[X_{i}\right]=\mu$ and $\operatorname{Var}(X)=\sigma^{2}<\infty$. If $\bar{X}_{n}=\frac{1}{m} \sum_{i=1}^{n} X_{i}$ then it holds that

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \mathbb{P}\left[\frac{\bar{X}_{m}-\mu}{\sigma / \sqrt{m}} \leq x\right]=\int_{-\infty}^{x} \frac{e^{-y^{2} / 2}}{\sqrt{2 \pi}} d y \quad \text { for all } x \in \mathbb{R} \tag{2.2}
\end{equation*}
$$

Hence, the central limit theorem states that if $m$ is large enough, the random variable $\frac{\bar{X}_{m}-\mu}{\sigma / \sqrt{m}}$ can be "approximated" by a standard normal random variable with distribution function $N(x)$, that is

$$
\begin{equation*}
\mathbb{P}\left[\frac{\bar{X}_{m}-\mu}{\sigma / \sqrt{m}} \leq x\right] \approx N(x)=\int_{-\infty}^{x} \frac{e^{-y^{2} / 2}}{\sqrt{2 \pi}} d y \quad \text { for } x \in \mathbb{R} \tag{22.3}
\end{equation*}
$$

The central limit theorem is one of the important results in probability theory. It has several important applications, for example

- constructing confidence intervals for the sample mean.
- analytically approximate discrete distributions, such as the binomial distribution.

Let us now study the latter application.

## 3. The normal approximation of the binomial distribution

Recall that a random variable $N_{m}$ is binomially distributed with parameter $m$ and $p$ if

$$
\begin{equation*}
N_{m}=\sum_{i=1}^{m} X_{i} \tag{3,1}
\end{equation*}
$$

where $X_{1}, X_{2}, \ldots X_{n}$ are independent Bernoulli distributed random variables with $\mathbb{P}\left[X_{i}=1\right]=$ $p$ and $\mathbb{P}\left[X_{i}=0\right]=1-p$, that is

$$
X_{i}= \begin{cases}1 & \text { with probability } p  \tag{36.2}\\ 0 & \text { with probability } 1-p\end{cases}
$$

So for any integer $j \leq m$ the probability $\mathbb{P}\left[N_{m}=k\right]$ is given by

$$
\begin{equation*}
\mathbb{P}\left[N_{m}=j\right]=\binom{m}{j} p^{j}(1-p)^{n-j} \tag{36.3}
\end{equation*}
$$

and thus for integers $k \leq m$ we have

$$
\begin{equation*}
\mathbb{P}\left[N_{m} \leq k\right]=\sum_{j=0}^{k}\binom{m}{j} p^{j}(1-p)^{n-j} . \tag{3,4}
\end{equation*}
$$

Note that for large $m$ the number $\binom{m}{j}$ will be extremely big for some $j$ and therefore numerically truncated in standard math software. For example computing $\binom{54}{23}$ in MatLab renders the result

```
>> nchoosek(54,23)
Warning:Result may not be exact.Coefficient is greater than 10^15,
and is only good to }15\mathrm{ digits.
> In nchoosek at line 55
ans = 1.0859e+015
```

Thus, for large $m$ we can not compute $\mathbb{P}\left[N_{m} \leq k\right]$ with the formula (3[4]). However, from (3:1) we note that $N_{m}=\sum_{i=1}^{m} X_{i}$ where the $X_{i}$-s are independent and equality distributed with mean $\mathbb{E}\left[X_{i}\right]=p$ and finite variance variance $\operatorname{Var}\left(X_{i}\right)=p(1-p)$. We can therefore apply the central limit theorem (CLT) as follows. In the binomial distribution case we have $\mu=p, \sigma=\sqrt{p(1-p)}$ and $\bar{X}_{m}=\frac{N_{m}}{m}$. So the quantity $\frac{\bar{X}_{m}-\mu}{\sigma / \sqrt{m}}$ can be rewritten as

$$
\begin{equation*}
\frac{\bar{X}_{m}-\mu}{\sigma / \sqrt{m}}=\frac{N_{m} / m-p}{\sqrt{p(1-p)} / \sqrt{m}}=\frac{N_{m}-m p}{\sqrt{m} \sqrt{p(1-p)}}=\frac{N_{m}-m p}{\sqrt{m p(1-p)}} \tag{3,5}
\end{equation*}
$$

and by the CLT we have that the random variable

$$
\begin{equation*}
\frac{\bar{X}_{m}-\mu}{\sigma / \sqrt{m}}=\frac{N_{m}-m p}{\sqrt{m p(1-p)}} \tag{3,6}
\end{equation*}
$$

is distributed as a standard normal random variable when $m$ is large enough. Hence, if $N_{m}$ is binomially distributed with parameter $m$ and $p$ we have for any integer $k \leq m$ that

$$
\begin{equation*}
\mathbb{P}\left[N_{m} \leq k\right]=\mathbb{P}\left[\frac{N_{m}-m p}{\sqrt{m p(1-p)}} \leq \frac{k-m p}{\sqrt{m p(1-p)}}\right] \approx N\left(\frac{k-m p}{\sqrt{m p(1-p)}}\right) \tag{3,7}
\end{equation*}
$$

where the last approximation is due to the central limit theorem when $m$ is large. Furthermore, as a rule of thumb the approximation works best if $m p(1-p)>5$ and thus becomes better for larger $m$. Hence, the central limit theorem implies the following result

Theorem 3.1. Normal approximation of the binomial distribution Let $N_{m}$ be binomially distributed with parameter $m$ and $p$. If $m p(1-p)>5$ then

$$
\begin{equation*}
\mathbb{P}\left[N_{m} \leq k\right] \approx N\left(\frac{k-m p}{\sqrt{m p(1-p)}}\right) \quad \text { for } k \leq m \tag{3,8}
\end{equation*}
$$

where $N(x)$ is the distribution function of a standard normal random variable.

By adding 0.5 to the $k$ in (318), the approximation of $\mathbb{P}\left[N_{m} \leq k\right]$ can often improve significantly. Hence, the quantity

$$
\begin{equation*}
N\left(\frac{k+0.5-m p}{\sqrt{m p(1-p)}}\right) \quad \text { for } k \leq m \tag{3,9}
\end{equation*}
$$

is often a better approximation to $\mathbb{P}\left[N_{m} \leq k\right]$ than the right hand side in (31.8). In Table 11 we clearly see that (319) consistently produce better approximations of $\mathbb{P}\left[N_{m} \leq k\right]$ for a binomially distributed random variable $N_{m}$ compared with the approximation (318).

Table 1. The distribution function $\mathbb{P}\left[N_{m} \leq k\right]$ for a binomially distributed random variable $N_{m}$ and the two versions (318) and (319) of the normal approximation of $\mathbb{P}\left[N_{m} \leq k\right]$, for different values of $(m, k, p)$. In all cases it holds that $m p(1-p)>5$.

| $(m, k, p)$ | $\mathbb{P}\left[N_{m} \leq k\right]$ | $N\left(\frac{k-m p}{\sqrt{m p(1-p)}}\right)$ | $N\left(\frac{k+0.5-m p}{\sqrt{m p(1-p)}}\right)$ |
| :---: | :---: | :---: | :---: |
| $(40,10,0.2)$ | 0.8392 | 0.7854 | 0.8385 |
| $(50,10,0.2)$ | 0.5836 | 0.5000 | 0.5702 |
| $(60,10,0.2)$ | 0.3234 | 0.2593 | 0.3141 |
| $(100,12,0.08)$ | 0.9441 | 0.9298 | 0.9514 |
| $(100,12,0.1)$ | 0.8018 | 0.7475 | 0.7977 |

4. The normal approximation of the mixed binomial distribution

Now consider a so called mixed binomial distribution. The mixed binomial distribution works as follows. Let $Z$ be a random variable on $\mathbb{R}$ with density $f_{Z}(z)$ and let $p(Z) \in[0,1]$ be a random variable with distribution $F(x)$ and mean $\bar{p}$, that is

$$
\begin{equation*}
F(x)=\mathbb{P}[p(Z) \leq x] \quad \text { and } \quad \mathbb{E}[p(Z)]=\int_{-\infty}^{\infty} p(z) f_{Z}(z) d z=\bar{p} \tag{4.1}
\end{equation*}
$$

Let $X_{1}, X_{2}, \ldots X_{m}$ be identically two-point distributed random variables in $\{0,1\}$, i.e. $X_{i}=$ 1 or $X_{i}=0$. Furthermore, conditional on $Z$, the random variables $X_{1}, X_{2}, \ldots X_{m}$ are independent and for $X_{i}=1$ with probability $p(Z)$, that is

$$
\begin{equation*}
\mathbb{P}\left[X_{i}=1 \mid Z\right]=p(Z) \tag{4.2}
\end{equation*}
$$

From rules of conditional probabilities we get that

$$
\mathbb{P}\left[X_{i}=1\right]=\mathbb{E}\left[X_{i}\right]=\mathbb{E}\left[\mathbb{E}\left[X_{i} \mid Z\right]\right]=\mathbb{E}[p(Z)]=\int_{-\infty}^{\infty} p(z) f_{Z}(z) d z=\bar{p}
$$

where the last equality is due to (4]1). Next, define $N_{m}$ as

$$
\begin{equation*}
N_{m}=\sum_{i=1}^{m} X_{i} . \tag{4.3}
\end{equation*}
$$

From the above model we conclude that conditionally on $Z$ the random variable $N_{m}$ will be a binomially distributed with parameter $m$ and $p(Z)$ so that

$$
\mathbb{P}\left[N_{m}=j \mid Z\right]=\binom{m}{j} p(Z)^{j}(1-p(Z))^{m-j}
$$

and since $\mathbb{P}\left[N_{m}=j\right]=\mathbb{E}\left[\mathbb{P}\left[N_{m}=j \mid Z\right]\right]=\mathbb{E}\left[\binom{m}{j} p(Z)^{j}(1-p(Z))^{j}\right]$ it holds that

$$
\begin{equation*}
\mathbb{P}\left[N_{m}=j\right]=\int_{-\infty}^{\infty}\binom{m}{j} p(z)^{j}(1-p(z))^{m-j} f_{Z}(z) d z \tag{4.4}
\end{equation*}
$$

Furthermore, as in the standard binomial model we then get that

$$
\begin{equation*}
\mathbb{P}\left[N_{m} \leq k\right]=\sum_{j=0}^{n}\binom{m}{j} \int_{-\infty}^{\infty} p(z)^{j}(1-p(z))^{m-j} f_{Z}(z) d z \tag{4.5}
\end{equation*}
$$

Note that computing $\mathbb{P}\left[N_{m} \leq k\right]$ in (4.5) suffers from the same problems and challenges as in the standard binomial model. However, just as in the standard binomial model we can exploit the central limit theorem as follows. So conditional on $Z$ the $X_{i}$-s are independent and equality distributed with conditional mean $\mathbb{E}\left[X_{i} \mid Z\right]=p(Z)$ and finite conditional variance $\operatorname{Var}\left(X_{i} \mid Z\right)=p(Z)(1-p(Z))$. Hence, conditional on $Z$ we can apply the central limit theorem. and then follow the same computations as in Section 3, especially Equation (31.7) then renders that

$$
\begin{align*}
\mathbb{P}\left[N_{m} \leq k \mid Z\right] & =\mathbb{P}\left[\frac{N_{m}-m p(Z)}{\sqrt{m p(Z)(1-p(Z))}} \leq \frac{k-m p(Z)}{\sqrt{m p(Z)(1-p(Z))}}\right]  \tag{4.6}\\
& \approx N\left(\frac{k-m p(Z)}{\sqrt{m p(Z)(1-p(Z)}}\right)
\end{align*}
$$

where of assume that $\mathbb{P}[0<p(Z)<1]=1$, that is $\mathbb{P}[p(Z)=1]=\mathbb{P}[p(Z)=0]=0$, so that the nominator is well defined. Next, noting that (416) implies

$$
\begin{equation*}
\mathbb{P}\left[N_{m} \leq k\right]=\mathbb{E}\left[\mathbb{P}\left[N_{m} \leq k \mid Z\right]\right]=\mathbb{E}\left[N\left(\frac{k-m p(Z)}{\sqrt{m p(Z)(1-p(Z)}}\right)\right] \tag{4.7}
\end{equation*}
$$

which gives that

$$
\begin{equation*}
\mathbb{P}\left[N_{m} \leq k\right] \approx \int_{-\infty}^{\infty} N\left(\frac{k-m p(z)}{\sqrt{m p(z)(1-p(z))}}\right) f_{Z}(z) d z \tag{4.8}
\end{equation*}
$$

where $f_{Z}(z)$ is the density of $Z$ as defined in Equation (4IT). Note that we here ignore the fact that the rule of thumb, $5>m p(z)(1-p(z))$, sometimes may be violated, and do not bother about this fact. So (4].8) is just a generalization of the normal approximation
formula (318) in Theorem [3.1. As already remarked, by adding 0.5 to the $k$ in (416), the approximation of $\mathbb{P}\left[N_{m} \leq k\right]$ can often improve significantly. Hence, the quantity

$$
\begin{equation*}
\int_{-\infty}^{\infty} N\left(\frac{k+0.5-m p(z)}{\sqrt{m p(z)(1-p(z)}}\right) f_{Z}(z) d z \quad \text { for } k \leq m \tag{4.9}
\end{equation*}
$$

is often a better approximation to $\mathbb{P}\left[N_{m} \leq k\right]$ than the right hand side in (4]8).
Note that given explicit expressions of the density $f_{Z}(z)$ and the probability $p(Z)$, then both formulas (4.8) and (4.9) are easy to evaluate with numerical quadrature, for example using quad in matlab.

## 5. The law of large numbers approximation of a mixed binomial DISTRIBUTION

Consider the same mixed binomial model as presented in Section 4 Then from the lecture slides in the "Financial risk"-course we know the following result.

Theorem 5.1. The approximation of the mixed binomial distribution using the law of large numbers With notation as above, let $N_{m}=\sum_{i=1}^{m} X_{i}$ be a mixed binomially distributed random variable where $\mathbb{P}\left[X_{i}=1 \mid Z\right]=p(Z)$. Then for any $x \in[0,1]$ it holds that

$$
\begin{equation*}
\mathbb{P}\left[\frac{N_{m}}{m} \leq x\right] \rightarrow \mathbb{P}[p(Z) \leq x]=F(x) \quad \text { as } \quad m \rightarrow \infty \tag{5,1}
\end{equation*}
$$

Thus, Theorem (5.1) implies that when $m$ is "large" we have the following approximation for any integer $k \leq m$

$$
\begin{equation*}
\mathbb{P}\left[N_{m} \leq k\right]=\mathbb{P}\left[\frac{N_{m}}{m} \leq \frac{k}{m}\right] \approx F\left(\frac{k}{m}\right) \tag{5,2}
\end{equation*}
$$

where where $F(x)=\mathbb{P}[p(Z) \leq x]$ for $x \in[0,1]$. Hence, besides the CLT-formulas (4]8) and (419), then Equation (512) is another alternative to approximate the probabilities $\mathbb{P}\left[N_{m} \leq k\right]$.

## 6. Finding the mixed binomial distribution by using Monte Carlo Simulations

From Section 4 and Section 5 we conclude that we have at least two (or three) formulas for approximating the mixed binomial distribution $\mathbb{P}\left[N_{m} \leq k\right]$, which are the Central limit (CLT) formulas (418) and (419), and the law of large number (LLN) formula (512). Both of these two (or three) formulas assumes that $m$ is large, and the following question now arise is: Which of the CLT and LLN approximations is the best approximation to the value $\mathbb{P}\left[N_{m} \leq k\right]$ ? In order to answer this question we must of course be able to compute the exact value of the probability $\mathbb{P}\left[N_{m} \leq k\right]$. However, as we have seen it is in general difficult to compute the exact distribution $\mathbb{P}\left[N_{m} \leq k\right]$ via the formula (47.5), which in fact was the motivation of using the CLT and LLN as approximations. Thus, how can we find such a close approximation of $\mathbb{P}\left[N_{m} \leq k\right]$ as possible? One possibility to find an accurate
estimate of this quantity is to use Monte Carlo simulation which simply means that we simulate $N_{m}$ many times (typically $10^{5}$ or more times) and for each such simulation check if $N_{m} \leq k$. By counting how many times the event $N_{m} \leq k$ will be true and then dividing with the number of simulations, we will by the law of large numbers get an estimate of the probability $\mathbb{P}\left[N_{m} \leq k\right]$. Let us formalize this argument more rigourously.

Computing $\mathbb{P}\left[N_{m} \leq k\right]$ using Monte-Carlo simulation

1. Simulate $n$ independent copies of $N_{m}$, that is $N_{m}^{(1)}, N_{m}^{(2)}, \ldots, N_{m}^{(n)}$.
2. Define $n$ random variables $Y_{1}^{(k)}, \ldots Y_{n}^{(k)}$ such that

$$
Y_{j}^{(k)}= \begin{cases}1 & \text { if } N_{m}^{(j)} \leq k  \tag{6,1}\\ 0 & \text { otherwise, i.e. if } N_{m}^{(j)}>k\end{cases}
$$

3. Compute $\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}$
4. Let $\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}$ be an estimate of the probability $\mathbb{P}\left[N_{m} \leq k\right]$.

Now, let us motivate why this so called Monte-Carlo simulation algorithm will yield an approximate of $\mathbb{P}\left[N_{m} \leq k\right]$. as the number of simulations $n$ increases (typically $n$ is given by e.g. $n=10^{3}, 10^{4}, 10^{5}, 10^{6}$ etc).
For notational convenience let $p_{m}^{(k)}=\mathbb{P}\left[N_{m} \leq k\right]$. Note that by construction we have that $\mathbb{E}\left[Y_{j}^{(k)}\right]=\mathbb{P}\left[N_{m} \leq k\right]=p_{m}^{(k)}$ because

$$
\mathbb{E}\left[Y_{j}^{(k)}\right]=1 \cdot p_{m}^{(k)}+0 \cdot\left(1-p_{m}^{(k)}\right)=p_{m}^{(k)}=\mathbb{P}\left[N_{m} \leq k\right] .
$$

Furthermore, since $Y_{j}^{(k)}$ is a two-point distributed random variable on $\{0,1\}$ we have that

$$
\begin{equation*}
\operatorname{Var}\left(Y_{j}^{(k)}\right)=p_{m}^{(k)}\left(1-p_{m}^{(k)}\right) \tag{6,2}
\end{equation*}
$$

Next, pick an arbitrary $\varepsilon>0$. Since $Y_{1}^{(k)}, \ldots Y_{n}^{(k)}$ are independent and equally distributed we can apply Chebyshev's inequality together with Equation (66.2) and retrieve

$$
\begin{aligned}
\mathbb{P}\left[\left|\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}-p_{m}^{(k)}\right| \geq \varepsilon\right] & \leq \frac{\operatorname{Var}\left(\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}\right)}{\varepsilon^{2}} \\
& =\frac{\frac{1}{n^{2}} \operatorname{Var}\left(\sum_{j=1}^{n} Y_{j}^{(k)}\right)}{\varepsilon^{2}} \\
& =\frac{\frac{1}{n^{2}} n \operatorname{Var}\left(Y_{j}^{(k)}\right)}{\varepsilon^{2}} \\
& =\frac{p_{m}^{(k)}\left(1-p_{m}^{(k)}\right)}{n \varepsilon^{2}}
\end{aligned}
$$

that is,

$$
\mathbb{P}\left[\left|\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}-p_{m}^{(k)}\right| \geq \varepsilon\right] \leq \frac{p_{m}^{(k)}\left(1-p_{m}^{(k)}\right)}{n \varepsilon^{2}}
$$

and we conclude that $\mathbb{P}\left[\left|\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}-p_{m}^{(k)}\right| \geq \varepsilon\right] \rightarrow 0$ as $n \rightarrow \infty$. Note that this holds for any $\varepsilon>0$. Hence, this is just a version the law of large numbers applied to the estimate $\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}$ of the probability $p_{m}^{(k)}=\mathbb{P}\left[N_{m} \leq k\right]$ as $n \rightarrow \infty$. So the random variable $\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}$ will converge in probability towards the constant $p_{m}^{(k)}=\mathbb{P}\left[N_{m} \leq k\right]$ as $n \rightarrow \infty$. In practice this means that for large values of $n$, the simulated random variable $\frac{1}{n} \sum_{j=1}^{n} Y_{j}^{(k)}$ will be (very) close to the value $p_{m}^{(k)}=\mathbb{P}\left[N_{m} \leq k\right]$ which is what we wanted to compute.

Finally, we here remark that step 1 in the above Monte-Carlo simulating algorithm is in turn split into the following steps.

## Monte-Carlo simulation of $X_{1}, \ldots, X_{m}$ and $N_{m}^{(j)}$

For each $j=1,2, \ldots, n$ do the following:
1.1. Simulate the random variable $Z$ and compute $p(Z) \in[0,1]$.
1.2. Simulate the i.i.d sequence $U_{1}, U_{2}, \ldots, U_{m}$ where $U_{i}$ is uniformly distributed on $[0,1]$ and independent of $Z$.
1.3. For each $i=1,2, \ldots, m$ define $X_{i}$ as

$$
X_{i}= \begin{cases}1 & \text { if } U_{i} \leq p(Z)  \tag{6.3}\\ 0 & \text { otherwise, i.e. if } U_{i}>p(Z)\end{cases}
$$

1.4. Compute $N_{m}^{(j)}=\sum_{i=1}^{m} X_{i}$.

Let us motivate why the above algorithm for generating the random variables $X_{1}, \ldots, X_{m}$ implies that $\mathbb{P}\left[X_{i}=1 \mid Z\right]=p(Z)$ for each $i=1,2, \ldots, m$. Let $F_{U_{i}}(x)=x$ be the distribution function for $U_{i}$ which is uniformly distributed on $[0,1]$. Given $p(Z)$ we then have by construction that

$$
\begin{equation*}
\mathbb{P}\left[X_{i}=1 \mid Z\right]=\mathbb{P}\left[U_{i} \leq p(Z) \mid Z\right]=F_{U_{i}}(p(Z))=p(Z) \tag{6.4}
\end{equation*}
$$

where the second equality is due to the fact that $U_{i}$ is independent of $Z$ and the equality follows from the observation $F_{U_{i}}(x)=x$ since $U_{i}$ is uniformly distributed on $[0,1]$.
6.1. Simulating $X_{1}, \ldots, X_{m}$ in the mixed Merton binomial model. Consider the mixed Merton binomial model. We know want to simulate the $m$ Bernoulli random variables $X_{1}, \ldots, X_{m}$ for a simulated value $Z$. In the mixed Merton binomial model $Z$ is a standard normal random variable and $p(Z)$ is given by

$$
\begin{equation*}
p(Z)=N\left(\frac{N^{-1}(\bar{p})-\sqrt{\rho} Z}{\sqrt{1-\rho}}\right) \tag{6.1.1}
\end{equation*}
$$

where $N(x)$ is the distribution function of a standard normal distribution and $\bar{p}=\mathbb{P}\left[X_{i}=1\right]$ while $\rho$ is the correlation parameters.

So simulating $X_{1}, \ldots, X_{m}$ can of course be done with the above algorithm in the steps 1.1 to step 1.4. However in the mixed Merton binomial model there is an alternative, and maybe more straightforward way to simulate $X_{1}, \ldots, X_{m}$ given $Z$. To see this, recall from the slides of the second credit risk lecture that (see slide 7 in lecture 2 (credit risk)),

$$
\begin{equation*}
X_{i}=1 \quad \text { is equivalent with } \quad Y_{i}<\frac{N^{-1}(\bar{p})-\sqrt{\rho} Z}{\sqrt{1-\rho}} \tag{6.1,2}
\end{equation*}
$$

where $Y_{1}, Y_{2}, \ldots, Y_{m}$ are independent and equally distributed standard normal random variables. Then we have the following alternative algorithm to steps 1.1 to step 1.4 above.

## Alternative Monte-Carlo simulation of $X_{1}, \ldots, X_{m}$ and $N_{m}^{(j)}$ in the mixed Merton binomial model

For each $j=1,2, \ldots, n$ do the following:
A1.1. Simulate the standard random variable $Z$.
A1.2. Simulate the independent and equally distributed standard normal random variables $Y_{1}, Y_{2}, \ldots, Y_{m}$ all independent of $Z$
A1.3. For each $i=1,2, \ldots, m$ define $X_{i}$ as

$$
X_{i}= \begin{cases}1 & \text { if } Y_{i} \leq \frac{N^{-1}(\bar{p})-\sqrt{\rho} Z}{\sqrt{1-\rho}}  \tag{6.11.3}\\ 0 & \text { otherwise, i.e. if } Y_{i}>\frac{N^{-1}(\bar{p})-\sqrt{\rho} Z}{\sqrt{1-\rho}}\end{cases}
$$

A1.4. Compute $N_{m}^{(j)}=\sum_{i=1}^{m} X_{i}$.
The above algorithm is thus an alternative to the algorithm in steps 1.1 to step 1.4 above, where we now avoid computing $p(Z)$ in (6.1.1).

