Computing the moment generating function:

$$\begin{aligned}
E[e^{tZ}] &= \sum_{i=1}^{m} e^{tZ_i}\Pr(Z = z_i) \\
&= 1 + tE[Z] + \sum_{j=1}^{m} \Pr(Z = z_i) \sum_{j=2}^{\infty} \frac{t^j z_i^j}{j!} \\
&= 1 + t^2 \sum_{i=1}^{m} \Pr(Z = z_i) z_i^2 \sum_{j=0}^{\infty} \frac{t^j z_i^j}{(j+2)!} \\
&\leq 1 + t^2 \sum_{i=1}^{m} \Pr(Z = z_i) z_i^2 \sum_{j=0}^{\infty} \frac{t^j M^j}{(j+2)!} \\
&\leq 1 + t^2 \frac{1}{2} \sum_{i=1}^{m} \Pr(Z = z_i) z_i^2 \sum_{j=0}^{\infty} \frac{t^j M^j}{j!} \\
&= 1 + \frac{e^{tM} t^2}{2} \sum_{i=1}^{m} \Pr(Z = z_i) z_i^2 \\
&= 1 + \frac{e^{tM} t^2}{2} E[Z^2] \\
&= 1 + \frac{e^{tM} t^2}{2} \mathrm{Var}(Z)
\end{aligned}$$

The justification for the second inequality is the fact that $(j+2)! \geq j!2!$. As a matter of fact, $(a+b)! \geq a!b!\ \forall a, b \geq 0$.

Lemma: With z as above, $E(e^{tz}) \leq 1 + \frac{e^{tM} t^2}{2} \mathrm{Var}(Z)$.

Proof of the Chernoff Bound

Let $|y_i| \leq M$. Now $y = \sum y_i$, $E[y_i] = M_i\ \forall i, \mathrm{Var}(y_i) = \sigma_i^2$ and the y_is are mutually independent. $E[Y] = \sum E[y_i] = \sum \mu_i$. $\mathrm{Var}(Y) = \mathrm{Var}(\sum y_i) = \sum \sigma_i^2$.

$$\begin{aligned}
\Pr(Y-\mu \geq \lambda_\sigma) &= \Pr(e^{t(Y-\mu)} \geq e^{t\lambda_\sigma}) \\
&\leq \frac{E[e^{t(Y-\mu)}]}{e^{t\lambda_\sigma}} \\
&= \frac{E[\prod_i e^{t(y_i-\mu_i)}]}{e^{t\lambda_\sigma}} \\
&= \frac{\prod_i [e^{t(y_i-\mu_i)}]}{e^{t\lambda_\sigma}} \\
&\leq \frac{\prod_i \left(1+\frac{t^2 e^{tM}}{2}\mathrm{Var}(y_i-\mu_i)\right)}{e^{t\lambda_\sigma}} \\
&\leq \frac{\prod_i \exp\left(t^2 \frac{e^{tM}}{2}\mathrm{Var}(y_i-\mu_i)\right)}{\exp(t\lambda_\sigma)} \\
&= \frac{\exp\left(t^2 \frac{e^{tM}}{2}\sum \sigma_i^2\right)}{\exp(t\lambda_\sigma)} \\
&= \exp\left(t^2 \frac{e^{tM}}{2}\sigma^2 - t\lambda_\sigma\right)
\end{aligned}$$

Then if $t \leq \dfrac{\ln 2}{M}$ we have $\dfrac{e^{tM}}{2} = 1$. Which gives $\exp(t^2\left(\dfrac{e^{tM}}{2}\right)\sigma^2 - t\lambda_\sigma) \leq \exp(t^2\sigma^2 - t\lambda_\sigma)$. The optimum occurs when

$$t = \frac{\lambda_\sigma}{2\sigma^2} = \frac{\lambda}{2\sigma}$$

If $\dfrac{\lambda}{2\sigma} \leq \dfrac{\ln 2}{M}$ then the choice of t is okay.

$\Pr(Y-\mu \geq \lambda_\sigma) \leq \exp(t^2\sigma^2 - t\lambda_\sigma)$. So $\Pr(Y-\mu \geq \lambda_\sigma) \leq \exp((\dfrac{\lambda}{2\sigma})^2\sigma^2 - \dfrac{\lambda}{2\sigma}\lambda_\sigma) = \exp(-\dfrac{\lambda^2}{4})$.

Since $M = 1$ for us, this works for $\dfrac{\lambda}{2\sigma} \leq \ln 2$.

Theorem: Let $y_1, y_2, ..., y_i$ be (mutually) independent random variables with $E[y_i] = \mu_i, |y_i| \leq 1 \mathrm{Var}(y_i) = \sigma_i^2$ and $y = \sum_{i=1}^{r} y_i, E[y] = \mu = \sum \mu_i$, $\mathrm{Var}[Y] = \sigma^2 = \sum \sigma_i^2$. Then $\Pr\left(|y-\mu| \geq \lambda_\sigma\right) \leq 2\exp\left(-\dfrac{\lambda^2}{4}\right)$ for all $\lambda, 0 \leq \lambda \leq \sigma$.

Example: Let y_i be a set of Bernoulli (p) random variables with the following:

$$\Pr(y_i) = \begin{cases} 0 & 1-p \\ 1 & p \end{cases}$$

$$E[y_i] = p$$

$$\mathrm{Var}(y_i) = p - p^2 = p(1-p)$$

Consider $y_1, y_2, ..., y_n$, $\mu = np$, $\sigma^2 = np(1-p)$.

Then we have $\Pr\left(|\sum y_i - np| \geq \lambda\sqrt{np(1-p)}\right) \leq 2\exp\left(-\frac{\lambda^2}{4}\right)$ if $0 \leq \lambda \leq \sqrt{np(1-p)}$.

Example: Now fix p and let $n \to \infty$.

Let $\lambda = \ln n$

$\Pr\left(|\sum y_i - np| \geq \ln n\sqrt{np(1-p)}\right) \leq 2\exp\left(-\frac{(\ln n)^2}{4}\right) \to 0.$

So with high probability if you flip n coins, the number of heads will be in

$$\left(\frac{n}{2} - \ln n\sqrt{\frac{n}{4}}, \frac{n}{2} + \ln n\sqrt{\frac{n}{4}}\right) = \left(\frac{n}{2} - \omega\sqrt{n}, \frac{n}{2} + \omega\sqrt{n}\right)$$

Now a couple quick notes. $\lim\limits_{n\to\infty} \frac{n \ln n}{n} = \lim\limits_{n\to\infty} \frac{\ln n}{\sqrt{n}}$

And $\ln(\ln n) << \ln n << n^\alpha$ for $\alpha > 0$.

A reminder about $\omega(g)$ and $o(f)$ notation: $f = \omega(g)$ if $\lim\limits_{n\to\infty} \frac{f}{g} \to \infty$. This also means $o(f) = g$.

Random Bipartite Graphs

Let $(A, B; E)$ be a bipartite graph with $|A| = |B| = L$. Let $a \in A$ be adjacent to $b \in B$ with probability p. All points (a, b) form an edge independently.

Let $X \subseteq A$ and $Y \subseteq B$.

So the density of (A, B) is $d(A, B) = \frac{e(A,B)}{|A||B|} = \frac{e(A,B)}{L^2}$.

The density of (X, Y) is $d(X, Y) = \dfrac{e(X, Y)}{|X||Y|}$.

With $E[e(X, Y)] = \displaystyle\sum_{|\{(x,y):x\in X, y\in Y\}|} p = |X||Y|p$, $E[d(A, B)] = |A||B|p$, and $E[d(A, B)] = p = E[d(X, Y)]$.

We want to show that, for a fixed ϵ and a fixed p that

$$\underline{P} = \Pr(\exists X \subseteq A, \exists Y \subseteq B, |X| \geq \epsilon L, |Y| \geq \epsilon L, |d(X, Y) - p| > \epsilon) \to 0$$

as $L \to \infty$.

By Boole's inequality we have that $\Pr(\bigcup A_i) \leq \sum \Pr(A_i)$. Then for some fixed X and Y we know that $\underline{P} \leq \binom{|A|}{\geq \epsilon L}\binom{|B|}{\geq \epsilon L}\Pr(|d(X, Y) - p| > \epsilon) \leq 2^{|A|}2^{|B|}\Pr(\text{for a fixed } X, Y, \text{ that } |d(x, y) - p| > \epsilon)$.

Fix X and Y. We have $|X||Y|$ r.v.s. $\mu = |X||Y|p$, $\sigma^2 = |X||Y|p(1 - p)$.

Each r.v. is ≤ 1 as it is Bernoulli.

$$\Pr(|e(X, Y) - |X||Y|p) \geq \lambda\sqrt{|X||Y|p(1 - p)} \leq 2\exp(-\frac{\lambda^2}{4})$$

for $0 \leq \lambda \leq \sqrt{|X||Y|p(1 - p)}$

$$\Pr(|d(X, Y) - p| \geq \lambda\sqrt{\frac{p(1 - p)}{|X||Y|}}) \leq 2\exp(-\frac{\lambda^2}{4})$$

for $0 \leq \lambda \leq \sqrt{|X||Y|p(1 - p)}$.

Let $\epsilon = \lambda\sqrt{\dfrac{p(1 - p)}{|X||Y|}}$ then $\lambda = \epsilon\sqrt{\dfrac{|X||Y|}{p(1 - p)}}$

We restrict $0 \leq \epsilon \leq p(1 - p)$

$\Pr\left(|d(X, Y) - p| \geq \epsilon\right) \leq 2\exp\left(-\dfrac{\epsilon^2}{4} \cdot \dfrac{|X||Y|}{p(1 - p)}\right) \leq 2\exp\left(-\dfrac{\epsilon^2}{4} \cdot \dfrac{\epsilon^2 L^2}{p(1 - p)}\right)$

So

$$4^L \Pr\left(\text{fixed } X, Y : d(X, Y) - p| > \epsilon\right) \leq \exp\left(L\ln 4 - \frac{\epsilon^4}{4p(1 - p)}L^2\right) \to 0$$

as $L \to \infty$.

If L is large enough all $X \subseteq A, Y \subseteq B, |X| \geq \epsilon L, |y| \geq \epsilon L$ will have density in $(d(A, B) - \epsilon, d(A, B) + \epsilon)$ with as high probability as we want.