Solutions for 5.3 to 5.5
Baiyu Li

March 5, 2015

Problem (5.3). Suppose there are m balls which are thrown into n bins, $m \geq n$. Prove that

1. If $m \geq c_1\sqrt{n}$ for constant c_1, then the probability of no two balls are in the same bin is at most $1/e$;
2. If $m \leq c_2\sqrt{n}$ for constant c_2, then the probability of no two balls are in the same bin is at least $1/2$.

Proof. Let E_i be the event that the i-th ball is thrown in an empty bin given that all previous balls are in distinct bins. Then $\Pr(E_i) = (1-(i-1)/n)$. Let E be the event that all balls are in distinct bins, then we have

$$\begin{aligned}
\Pr(E) &= \Pr(E_1)\Pr(E_2)\cdots\Pr(E_m) \\
&= \Big(1-\frac{1}{n}\Big)\cdots\Big(1-\frac{m-1}{n}\Big) \\
&= \prod_{i=1}^{m}\Big(1-\frac{i-1}{n}\Big) \\
&\leq \prod_{i=1}^{m} e^{-\frac{i-1}{n}} \\
&= e^{-\sum_{i=1}^{m}\frac{i-1}{n}} \\
&= e^{-\frac{m(m-1)}{2n}}
\end{aligned}$$

Consider part (1) first. Since $m-1 \geq m/2$, we have $m(m-1)/2n \geq m^2/4n$. Let $c_1 = 2$ and suppose $m \geq 2\sqrt{n}$. Then $m(m-1)/2n \geq m^2/4n \geq 1$ and hence $e^{-\frac{m(m-1)}{2n}} \leq 1/e$.

Now for part (2), we can lower bound $\Pr(E)$ by assuming $m \leq n/2$. Then

$$\begin{aligned}
\Pr(E) &= \Big(1-\frac{1}{n}\Big)\cdots\Big(1-\frac{m-1}{n}\Big) \\
&= \prod_{i=1}^{m}\Big(1-\frac{i-1}{n}\Big) \\
&\geq e^{-\sum_{i=1}^{m}(\frac{i-1}{n}-\frac{(i-1)^2}{n^2})}.
\end{aligned}$$

Note that

$$\sum_{i=1}^{m}\Big(\frac{i-1}{n}-\frac{(i-1)^2}{n^2}\Big) \leq \frac{m^2}{2n}+\frac{m^3}{3n^2} \leq \frac{m^2}{2n}+\frac{m^2}{6n} = \frac{2m^2}{3n}.$$

When $m \leq \sqrt{\frac{3\ln 2}{2}n} \leq n/2$ for large n, we have $\Pr(E) \geq e^{-\frac{2m^2}{3n}} \geq 1/2$. $\square$

Problem (5.4). Suppose a room has 100 people. What is the probability of having three people sharing the same birthday?

Proof. Let $m = 100$, $n = 365$. Let X be the random variable representing the number of pairs of people with the same birthday. We have that

$$\begin{aligned}\Pr(X = r) &= \binom{n}{r}\left(\frac{1}{n}\right)^{2r}\binom{m}{2}\binom{m-2}{2}\cdots\binom{m-2r+2}{2}\prod_{i=m-2r+1}^{m}\left(1 - \frac{i-r-1}{n}\right) \\ &= \binom{n}{r}\left(\frac{1}{n}\right)^{2r}\frac{m!}{2^r(m-2r)!}\prod_{i=m-2r+1}^{m}(1 - \frac{i-r-1}{n})\end{aligned}$$

Then the probability of not having three people sharing the same birthday is

$$\sum_{r=0}^{m/2}\Pr(X = r) = \sum_{r=0}^{m/2}\binom{n}{r}\left(\frac{1}{n}\right)^{2r}\frac{m!}{2^r(m-2r)!}\prod_{i=m-2r+1}^{m}(1 - \frac{i-r-1}{n})$$

where $m = 100$ and $n = 365$. ☐

Problem (5.5). Suppose X is a Poisson random variable with mean μ where X represents the number of errors in a book. For each error, the probability of being a grammatical error is p and the probability of being a spelling error is $1 - p$. Let Y and Z be random variables representing the number of grammatical errors and spelling errors. Show that Y and Z are independent Poisson random variables with means $p\mu$ and $(1 - p)\mu$.

Proof. We know that $\Pr(Y = j) = \sum_i \Pr(X = i, Y = j) = \sum_i \Pr(Y = j|X = i)\Pr(X = i)$. For $i \geq j$, we have

$$\Pr(Y = j|X = i) = \binom{i}{j}p^j(1-p)^{i-j},$$

and for $i < j$, $\Pr(Y = j|X = i) = 0$. Then

$$\begin{aligned}\Pr(Y = j) &= \sum_{i=j}^{\infty}\binom{i}{j}p^j(1-p)^{i-j}\frac{e^{-\mu}\mu^i}{i!} \\ &= \sum_{k=0}^{\infty}\binom{j+k}{j}p^j(1-p)^k\frac{e^{-\mu}\mu^{j+k}}{k!} \\ &= \frac{p^j\mu^j}{j!}e^{-p\mu}\sum_{k=0}^{\infty}\frac{((1-p)\mu)^k e^{-(1-p)\mu}}{k!} \\ &= \frac{p^j\mu^j}{j!}e^{-p\mu}.\end{aligned}$$

Hence Y is a Poisson random variable with mean $p\mu$. Similarly, Z is also a Poisson random variable with mean $(1 - p)\mu$.

To show that Y and Z are independent, consider the joint distribution $\Pr(Y = i, Z = j)$. Since each error can be either a grammatical or a spelling error, we have $\Pr(Y = i, Z = j) = \Pr(Y = i, X = i + j)$, and hence

$$\begin{aligned}
\Pr(Y=i, Z=j) &= \Pr(Y=i, X=i+j) \\
&= \Pr(Y=i|X=i+j)\Pr(X=i+j) \\
&= \binom{i+j}{i} p^i (1-p)^j \frac{e^{-\mu}\mu^{i+j}}{(i+j)!} \\
&= \frac{p^i(1-p)^j}{i!j!} e^{-\mu}\mu^{i+j} \\
&= \Pr(Y=i)\Pr(Z=j)
\end{aligned}$$

Therefore Y and Z are independent. □