MTH135/STA104: Probability

Homework # 3 Due: Tuesday, Sep 20, 2005

Prof. Robert Wolpert

1. (from prob 7 *p.* 91) You roll a (fair, six-sided) die and I roll a die. You win if the number showing on your die is strictly greater than the one on mine.

a) What is the probability that you win our first game?

b) If we play the game five times, what is the probability that you win at least four times?

2. A sample of size $k = 10$ is taken with replacement from a population of size n. What is the smallest value of n so that the chance of a duplicate (a single individual being chosen two or more times) is below 50%?

Need

$$0.50 \le f(n) \equiv \mathsf{P}[\text{no duplicate}|n] = \frac{n(n-1)\cdots(n-9)}{n^{10}}$$

Evidently

$$\begin{aligned} f(68) = \frac{68\cdot 67\cdot 66\cdot 65\cdot 64\cdot 63\cdot 62\cdot 61\cdot 60\cdot 59}{68^{10}} &\approx 0.499111056351397 \\ &< 0.5 < \\ f(69) = \frac{69\cdot 68\cdot 67\cdot 66\cdot 65\cdot 64\cdot 63\cdot 62\cdot 61\cdot 60}{69^{10}} &\approx 0.504420222405075, \end{aligned}$$

so the answer is 69. For more methodical approach, use Stirling's approximation:

$$\begin{aligned}
f(n) &= \frac{n!\,n^{-10}}{(n-10)!} \\
&\approx \frac{\sqrt{2\pi n}\,n^n\,e^{-n}\,n^{-10}}{\sqrt{2\pi(n-10)}\,(n-10)^{n-10}\,e^{-(n-10)}} \\
&= (1-10/n)^{9.5-n}\,e^{-10}, \text{ so} \\
\log f(n) &= (9.5-n)(-10/n-100/2n^2+o(1/n^3))-10 \\
&= -95/n+10+50/n-10+o(1/n^2) \\
&\approx -45/n+o(1/n^2),
\end{aligned}$$

so we need n just big enough that $45/n > \log(2) \approx 0.6931$, *i.e.*, $n \geq 45/\log(2) \approx 64.92$. A little fiddling shows that 69 is the exact answer.

3. (prob 8 *p.* 91) For each positive integer n, find the largest value of p such that zero is the most likely number of successes in n independent trials, each with probability p of success.

We want $(1-p)^n = P(0) \geq P(1) = n\,p(1-p)^{n-1}$, so $(1-p) \geq n\,p$ or $1 \geq p(n+1)$ and hence $p = 1/(n+1)$ is the largest p that works.

4. 75% of the people in a certain population are left-handed. A random sample of size 15 will be drawn, with replacement, from this population.

a) What is the most likely number of left-handed persons in the sample?

The most likely number k satisfies $np-q \leq k \leq np+p$; here $p=0.75$ and $n=15$, so $11.25-0.25 \leq k \leq 11.25+0.75$, *i.e.*, $k=11$ or 12.

b) What is the probability of getting exactly this number of left-handed people in the sample?

The probability of each of these is $P(12) = \binom{15}{12}(0.75)^{12}0.25^3 = 455\,3^{12}/4^{15} \approx 0.2251991$.

5. Use geometry and art skills to evaluate the following integrals. Show your pictures.

a) $\int_{-2}^{2}\int_{0}^{\sqrt{4-x^2}} 1\;dy\;dx \quad = \pi 2^2/2 = 2\pi$

b) $\int_{-2}^{2}\int_{0}^{|2-x|} 1\;dy\;dx \quad = \frac{1}{2}\cdot 4\cdot 2 = 4$

c) $\int_0^1 x\sqrt{3}\;dx + \int_1^2 \sqrt{4-x^2}\;dx \quad = \pi 2^2/6 = 2\pi/3$

6. Approximately 52% of human live births are male in the U.S. Let M be the number of males among 400 live births at Durham Regional Hospital; assume the childrens' genders are independent. Find normal approximations to four correct decimals to

$$
\begin{array}{llll}
\text{a)}\quad \mathsf{P}[190 \le M \le 210] & \approx \Phi\left(\frac{210.5-208}{\sqrt{99.84}}\right) & -\Phi\left(\frac{189.5-208}{\sqrt{99.84}}\right) & \\
& = \Phi(0.2502) & -\Phi(-1.8515) & \\
& \approx 0.5988 & -0.0321 & \approx 0.5667 \\
\text{b)}\quad \mathsf{P}[210 \le M \le 220] & \approx \Phi\left(\frac{220.5-208}{\sqrt{99.84}}\right) & -\Phi\left(\frac{209.5-208}{\sqrt{99.84}}\right) & \\
& = \Phi(1.251) & -\Phi(0.1501) & \\
& \approx 0.8945 & -0.5597 & \approx 0.3349 \\
\text{c)}\quad \mathsf{P}[220 \le M \le 230] & \approx \Phi\left(\frac{230.5-208}{\sqrt{99.84}}\right) & -\Phi\left(\frac{219.5-208}{\sqrt{99.84}}\right) & \\
& = \Phi(2.2518) & -\Phi(1.1509) & \\
& \approx 0.9878 & -0.8751 & \approx 0.1127 \\
\text{d)}\quad \mathsf{P}[M = 200] & \approx \Phi\left(\frac{200.5-208}{\sqrt{99.84}}\right) & -\Phi\left(\frac{199.5-208}{\sqrt{99.84}}\right) & \\
& = \Phi(-0.7506) & -\Phi(-0.8507) & \approx 0.0290 \\
\qquad\qquad\qquad\text{or} & \approx \varphi\left(\frac{200-208}{\sqrt{99.84}}\right)/\sqrt{99.84} & & \approx 0.02898 \\
\text{e)}\quad \mathsf{P}[M = 210] & \approx \Phi\left(\frac{210.5-208}{\sqrt{99.84}}\right) & -\Phi\left(\frac{209.5-208}{\sqrt{99.84}}\right) & \\
& = \Phi(0.2502) & -\Phi(0.1501) & \approx 0.0391 \\
\qquad\qquad\qquad\text{or} & \approx \varphi\left(\frac{210-208}{\sqrt{99.84}}\right)/\sqrt{99.84} & & \approx 0.03913
\end{array}
$$

7. (prob 9, *p.* 109) An airline knows that over the long run 90% of passengers who reserve seats show up for their flight. On a particular flight with 300 seats, the airline accepts 324 reservations.

a) Assuming that passengers show up independently, what is the chance that the plane will be overbooked? Give answer as a decimal to four correct decimals.

$$
\begin{aligned}
\mathsf{P}[X > 300] &= \sum_{k=301}^{324} \binom{324}{k} 0.90^k 0.10^{324-k} \\
&\approx 1 - \Phi\left(\frac{300.5 - 324 \times 0.90}{\sqrt{324 \times 0.90 \times 0.10}}\right) \\
&= \Phi(-1.6481) \approx 0.04966
\end{aligned}
$$

b) Suppose that people tend to travel in groups. Would that decrease or increase the probability of overbooking? Explain your answer.

It would increase the chance of overbooking, up to a maximum of 90% if all 324 passengers flew together as a group— see below for illustration:

c) In particular, suppose the 324 reserved passengers were 162 couples. Now what is the chance of overbooking?

$$\begin{aligned} \mathsf{P}[X > 300] &= \sum_{k=151}^{162} \binom{162}{k} 0.90^k 0.10^{162-k} \\ &\approx 1 - \Phi\left(\frac{150.5 - 162 \times 0.90}{\sqrt{162 \times 0.90 \times 0.10}}\right) \\ &= \Phi(-1.2309) \approx 0.1092 \end{aligned}$$

8. A pollster asks a random sample of some number n of potential voters whether they prefer candidate A or candidate B; denote by x the number in her sample who prefer A. She wishes to estimate the proportion p of the *population* who favor A; her estimate will be the proportion x/n of the *sample* who favor A.

a) Find an expression in terms of n, p, and the Normal CDF function $\Phi(\cdot)$ for the (approximate) probability that her sample estimate x/n will *not* be within ± 0.02 (within two percent) of the true value p.

We use the normal approximation to the binomial, with $\mu = n\,p$ and

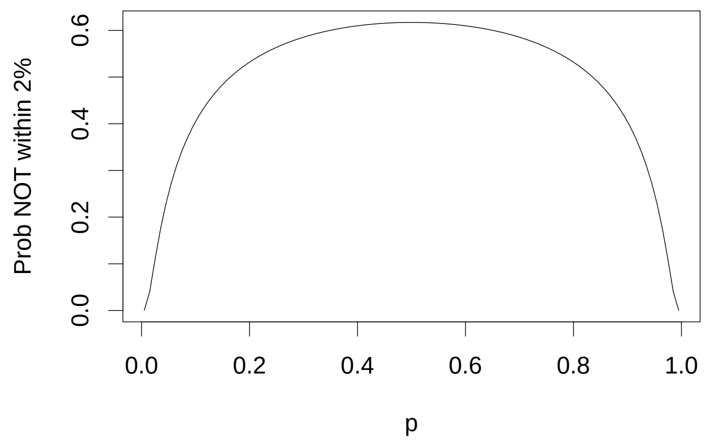
$\sigma^2 = n\,p\,q$, where as usual $q = 1 - p$:

$$\begin{aligned}
\mathsf{P}[\text{Sample } \textit{is} \text{ within } \pm 2\%] &= \mathsf{P}[-0.02 \le \frac{x}{n} - p \le +0.02] \\
&= \mathsf{P}[np - 0.02\,n \le x \le np + 0.02\,n] \\
&\approx \Phi\left(\frac{np + 0.02n + \frac{1}{2} - \mu}{\sigma}\right) - \Phi\left(\frac{np - 0.02n - \frac{1}{2} - \mu}{\sigma}\right) \\
&= \Phi\left(\frac{np + 0.02n + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{np - 0.02n - \frac{1}{2} - np}{\sqrt{npq}}\right) \\
&= \Phi\left(\frac{0.02n + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{-0.02n - \frac{1}{2}}{\sqrt{npq}}\right) \\
&= 2\Phi\left(\frac{0.02\sqrt{n} + \frac{1}{2\sqrt{n}}}{\sqrt{pq}}\right) - 1, \text{ so} \\
\mathsf{P}[\text{Sample is } \textit{not} \text{ within } \pm 2\%] &= 2 - 2\Phi\left(\frac{0.02\sqrt{n} + \frac{1}{2\sqrt{n}}}{\sqrt{pq}}\right) \\
&= 2\Phi\left(-\frac{0.02\sqrt{n} + \frac{1}{2\sqrt{n}}}{\sqrt{pq}}\right) \qquad (1)
\end{aligned}$$

Those who skipped the $\pm\frac{1}{2}$ will have the somewhat simpler expression $\Phi\left(-0.02\sqrt{n/pq}\right)$, which will be good enough for most purposes.

b) With $n = 100$, plot the error probability in your answer to part as a function of p for $0 < p < 1$.

For $n = 100$, Eqn. (1) becomes $2\Phi(-0.25/\sqrt{pq})$ and we get:



c) What is the *maximum* value of this error probability over the range $0 < p < 1$?

Evidently the function attains a maximum at $p = \frac{1}{2}$, where its value is $2\Phi(-0.5) \approx 0.6171$. Those who skipped the $\pm\frac{1}{2}$ will find $2\Phi(-0.4) \approx 0.6892$, probably close enough.

d) How large must n be for the maximum error probability to be no more than 5%? This is how the Gallup, Roper, and other polling companies choose their sample sizes.

The function of Eqn. (1) reaches its maximum at $p = \frac{1}{2}$, so we need

$$\begin{aligned}
0.05 = 2\Phi(-1.96) &\geq 2\Phi\left(-\frac{0.02\sqrt{n} + \frac{1}{2\sqrt{n}}}{\sqrt{\frac{1}{2}\cdot\frac{1}{2}}}\right) \\
&= 2\Phi\left(-0.04\sqrt{n} - 1/\sqrt{n}\right), \text{ so} \\
-1.96 &\geq -0.04\sqrt{n} - 1/\sqrt{n}, \text{ so } x \equiv \sqrt{n} \text{ satisfies the quadratic} \\
0.04x^2 - 1.96x + 1 &\geq 0 \\
\sqrt{n} = x &\geq \frac{1.96 + \sqrt{1.96^2 - 0.16}}{0.08} \approx 48.48437 \\
n &\geq 48.48437^2 = 2350.734,
\end{aligned}$$

so the sample would have to be about 2351 to achieve $\pm 2\%$ error with probability at least 95%.

Those who skipped the $+\frac{1}{2\sqrt{n}}$ in Eqn. (1) will have an easier time; their derivation will be

$$\begin{aligned} 0.05 = 2\Phi(-1.96) &\geq 2\Phi\left(-0.04\sqrt{n}\right), \text{ so } n \text{ satisfies} \\ -1.96 &\geq -0.04\sqrt{n} \text{ or} \\ n &\geq (1.96/0.04)^2 = 2401 \end{aligned}$$

so they'll think they need a little bigger sample.