Lecture 7

Repeated trials

Binomial distribution

"Normal" approximation

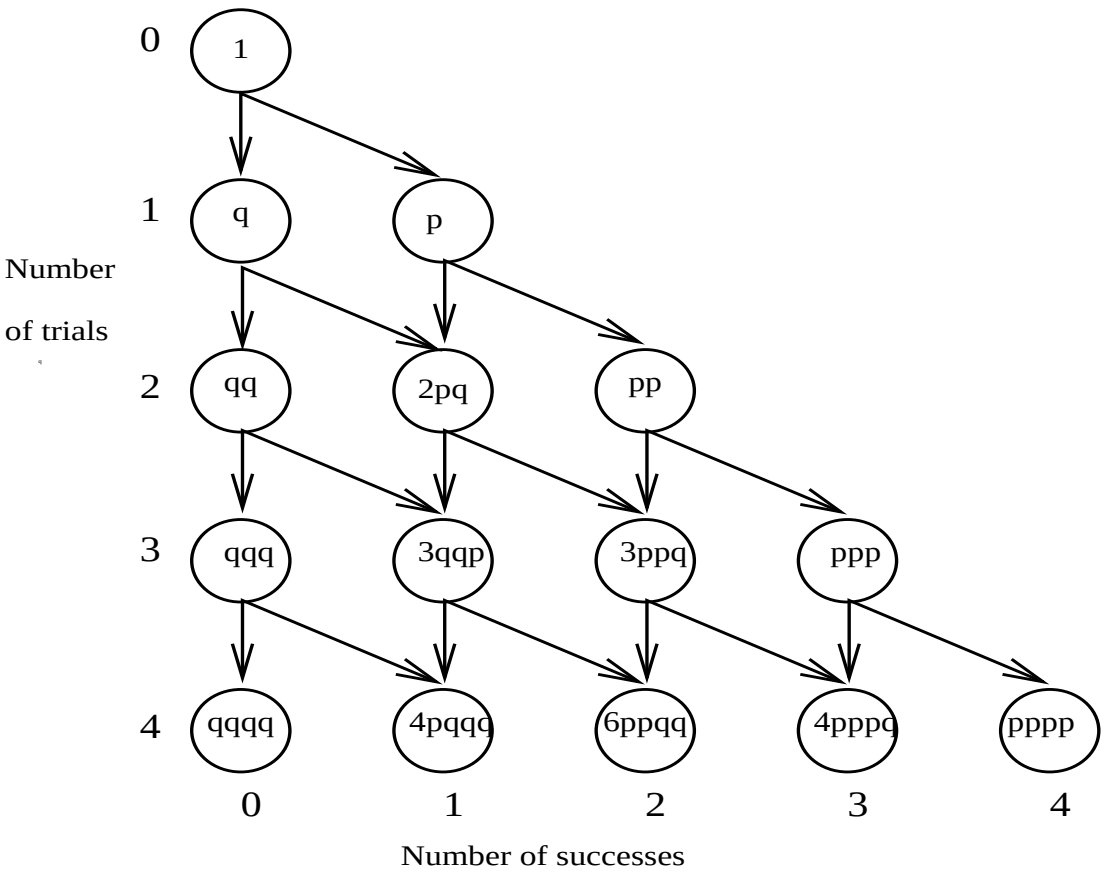
Poisson approximation

Poisson distribution

Success & Failure

Nature of trial	Success	Failure	p and q
Coin toss	Head	Tail	1/2 and 1/2
Rolling a die	6	Not 6	1/6 and 5/6
Rolling 2 dice	Sum of 7	Not a sum of 7	1/6 and 5/6
Birth of child	Son	Daughter	0.513 and 0.487
Probability exam	Pass	Fail	? and 1-?

Binomial tree



$$(p+q)^4 \;=\; q^4 + 4pq^3 + 6p^2q^2 + 4p^3q + p^4$$

Binomial distribution

k successes in n independent (Bernoulli) trials

- $P\{\underbrace{SS\ldots\ldots S}_{k \text{ times}}\ \underbrace{FF\ldots\ldots F}_{n-k \text{ times}}\} = p^k q^{n-k}$
- $P\{F\underbrace{SS\ldots\ldots S}_{k \text{ times}}\ \underbrace{FF\ldots\ldots F}_{n-k-1 \text{ times}}\} = qp^k q^{n-k-1} = p^k q^{n-k}$

$\Rightarrow$

- $P\{k \text{ success out of } n \text{ trials}\} = \binom{n}{k} p^k q^{n-k}, \qquad 0 \le k \le n.$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}$$

is the number of ways to choose k things out of n.

Binomial facts

X is the number of successes in n Bernoulli trials

- $P\{X = k\} = \binom{n}{k} p^k q^{n-k}, \qquad 0 \le k \le n.$
- $\binom{n}{k} \;=\; \binom{n-1}{k} + \binom{n-1}{k-1}.$

$$\begin{aligned}
\binom{n-1}{k} + \binom{n-1}{k-1} &= (n-1)! \left[\frac{1}{k!(n-1-k)!} + \frac{1}{(k-1)!(n-k)!} \right] \\
&= \frac{(n-1)!}{k!(n-k)!} \left[(n-k) + k \right] \\
&= \binom{n}{k}
\end{aligned}$$

- $(p+q)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \;=\; \sum_{k=0}^{n} P\{X = k\} \;=\; 1.$

- 1: $p = q = 1/2 \;\Rightarrow P\{X = k\} = \binom{n}{k} 2^{-n}$.
- 2: $A = \{\geq 4$ heads in 6 tosses of a fair coin.$\}$

$$\begin{aligned} P\{A\} &= P\{X=4\} + P\{X=5\} + P\{X=6\} \\ &= 2^{-6}\left[\binom{6}{4} + \binom{6}{5} + \binom{6}{6}\right] \\ &= [15+6+1]/64 \;=\; 11/32 \end{aligned}$$

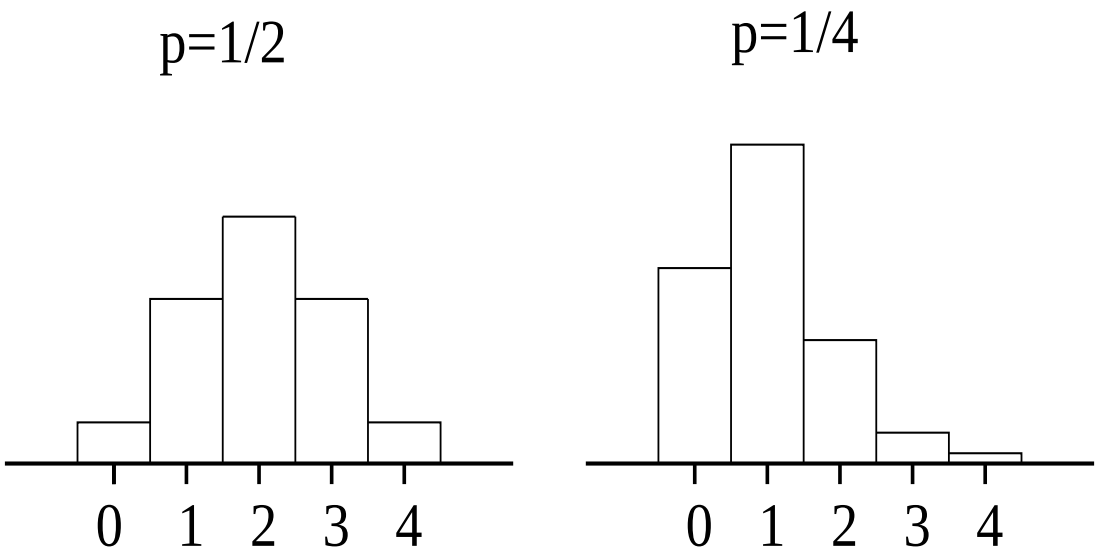
- 3: $A = \{$Among 5 families, each with 6 children, at least 3 families have at least 4 girls.$\}$

  - Assume that $P\{\text{girl}\} = P\{\text{boy}\} = 1/2$.
  - "Success" = "A family of 6 has at least 4 girls"$\Rightarrow p = 11/32$.

$$P\{A\} = P\{X=3\} + P\{X=4\} + P\{X=5\}$$
$$= \binom{5}{3}\left(\frac{11}{32}\right)^3\left(\frac{21}{32}\right)^2 + \binom{5}{4}\left(\frac{11}{32}\right)^4\left(\frac{21}{32}\right)^1 + \left(\frac{11}{32}\right)^5 = 0.226$$

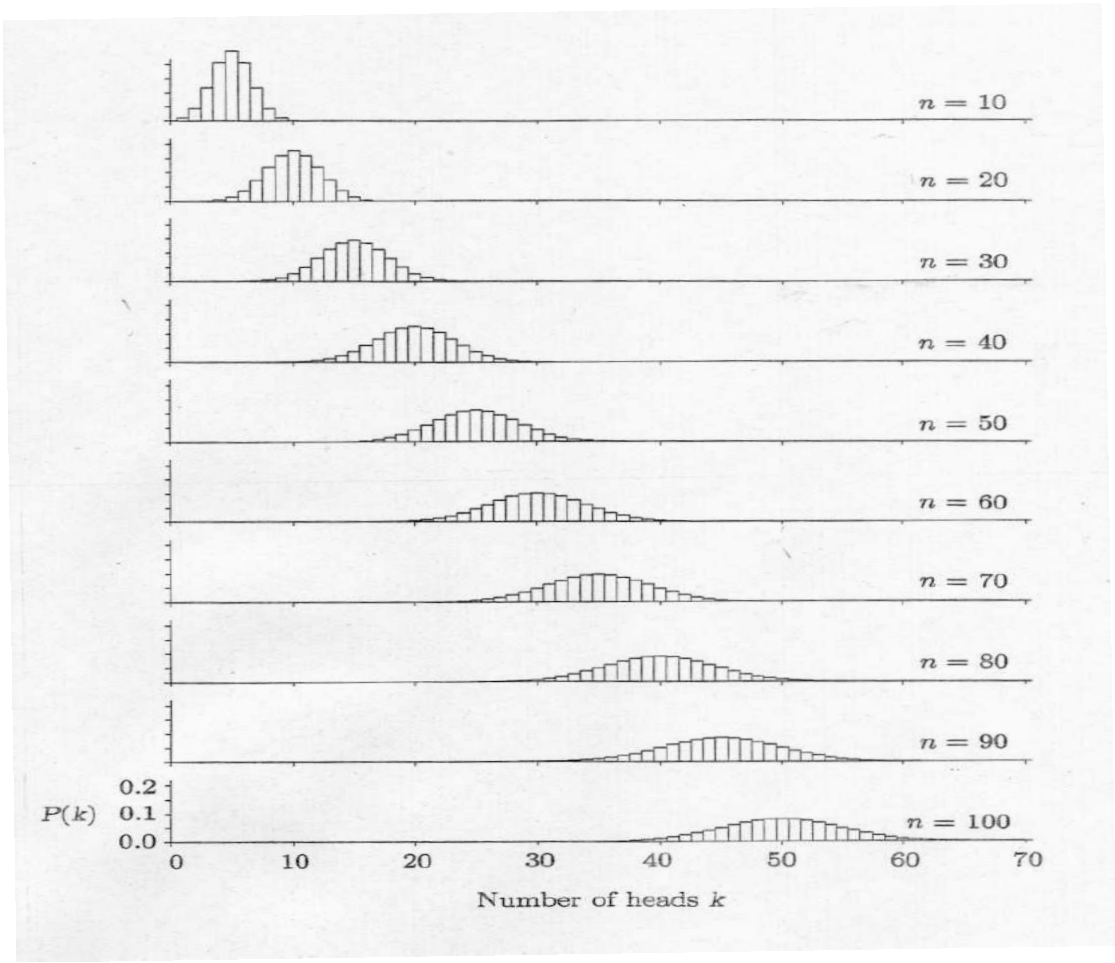
Binomial probabilities

$$\binom{n}{k} p^k q^{n-k}$$

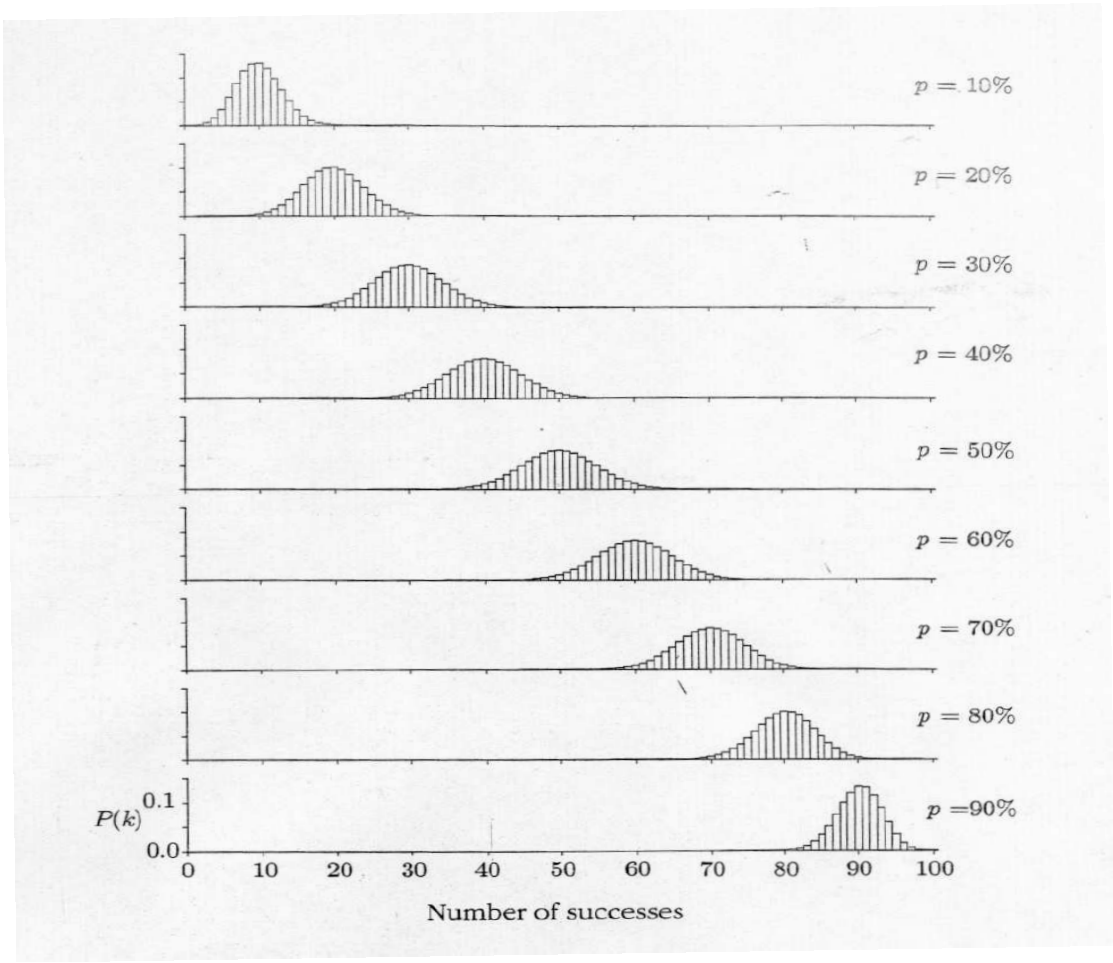


"Mean" $\approx$ "Mode" $\approx np$.

Binomial probabilities for $p = 0.5$ and varying n



Binomial probabilities for $n = 100$, and varying p



Mode of Bin(n, p) distribution

- We are looking for the <u>value</u> with the <u>largest</u> probability; i.e. the k^* for which

$$P\{k^* - 1\} \leq P\{k^*\} \quad \text{and} \quad P\{k^*\} \geq P\{k^* + 1\}.$$

$$\begin{aligned}
P\{k-1\} \leq P\{k\} \quad &\Longleftrightarrow \quad 1 \;\leq\; P\{k\}/P\{k-1\} \\
&\Longleftrightarrow \quad 1 \;\leq\; \binom{n}{k} p^k q^{n-k} \Big/ \binom{n}{k-1} p^{k-1} q^{n-k+1} \\
&\Longleftrightarrow \quad 1 \;\leq\; p/(1-p) \times (n-k+1)/k \\
&\Longleftrightarrow \quad k(1-p) \;\leq\; (n-k+1)p \\
&\Longleftrightarrow \quad k \;\leq\; (n+1)p
\end{aligned}$$

$\Rightarrow$ the mode k^* satisfies $k^* \leq np + p \leq k^* + 1$, or, $k^* = [np + p]$, the "largest integer less than or equal to $np + p$".

Computing probabilities

1: Recursively

$$X \sim \text{Bin}(n,p) \quad \Rightarrow \quad \frac{P\{k\}}{P\{k-1\}} = \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k-1} p^{k-1} q^{n-k+1}}$$

$$= \frac{p}{q}\left[\frac{n-k+1}{k}\right]$$

- Example: $n = 4,\ p = 0.9,\ \Rightarrow p/q = 9$

k	0	1	2	3	4
Formula	0.1^4	$9\frac{4}{1}P\{0\}$	$9\frac{3}{2}P\{1\}$	$9\frac{2}{3}P\{2\}$	$9\frac{1}{4}P\{3\}$
Probability	0.0001	0.0036	0.0486	0.2916	0.6561

2: The "normal" approximation

$X \sim \text{Bin}(n, p)$, n large (≥ 30) and p not "too far" from $1/2$ $\Rightarrow$

$$P\{a \leq X \leq b\} \approx \frac{1}{\sqrt{2\pi}} \int_{(a-0.5-\mu)/\sigma}^{(b+0.5-\mu)/\sigma} e^{-x^2/2}\, dx$$

where $\mu = np$ and $\sigma = \sqrt{npq}$. Accuracy $\approx \frac{|1-2p|}{10\sqrt{npq}}$.

3: The Poisson approximation

$X \sim \text{Bin}(n, p)$ with n large, p small, and np reasonable, $\Rightarrow$

$$P\{X = k\} \approx e^{-\mu} \frac{\mu^k}{k!},$$

where $\mu = np$.

$$\boxed{P\{X = k\} \;\approx\; e^{-\mu}\mu^k/k!.}$$

- Suppose that individual, healthy young students are totally incapacitated by illness, on average, 3 days a year. Suppose that there are 300 students taking an exam in probability. What is the probability that <u>more than 4</u> students are sick on the day of the exam?

- "Success" = sick on day of exam $\Rightarrow \quad p = 3/365 = 0.0082.$
- X = Number of sick students $\quad\Rightarrow\quad X \sim \text{Bin}(300,\ 3/365).$
- $\mu \;=\; np \;=\; 300 \times 3/365 \;=\; 2.466.$

$$P\{> 4 \text{ sick}\} \;=\; 1 - P\{\leq 4 \text{ sick}\}$$
$$= \; 1 - e^{-2.466}\left[\frac{2.466^0}{0!} + \frac{2.466^1}{1!} + \frac{2.466^2}{2!} + \frac{2.466^3}{3!} + \frac{2.466^4}{4!}\right]$$
$$= \; 0.1043$$

Same problem: No approximation

- "Success" = sick on day of exam $\Rightarrow \quad p = 3/365 = 0.0082.$
- X = Number of sick students $\Rightarrow \quad X \sim \text{Bin}(300,\ 3/365).$

$$
\begin{aligned}
P\{> 4 \text{ sick}\} &= 1 - P\{\leq 4 \text{ sick}\} \\
&= 1 - \left[\binom{300}{0} 0.0082^0 0.9918^{300} + \binom{300}{1} 0.0082^1 0.9918^{299}\right. \\
&\qquad + \binom{300}{2} 0.0082^2 0.9918^{298} + \binom{300}{3} 0.0082^3 0.9918^{297} \\
&\qquad \left. + \binom{300}{4} 0.0082^4 0.9918^{296}\right]
\end{aligned}
$$

which is difficult to compute.

Poisson approximation: Proof

$$
\begin{aligned}
P\{X = k\} &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)\ldots(n-k+1)}{k!} \times p^k (1-p)^{n-k} \\
&= n^k \times 1(1-1/n)\ldots(1-(k-1)/n) \\
&\qquad \times p^k (1-p)^n (1-p)^{-k}/k! \\
&= 1(1-1/n)\ldots(1-(k-1)/n) \times (1-p)^{-k} \\
&\qquad \times \mu^k (1-\mu/n)^n/k! \\
&\approx e^{-\mu}\mu^k/k!
\end{aligned}
$$

if p is small, k is fixed, $n \to \infty$, and $np \equiv \mu$.

Remember: $(1 + x/n)^n \to e^x$ as $n \to \infty$!

Poisson distribution

Take

$$p_k \equiv \frac{e^{-\mu}\mu^k}{k!}, \qquad k \geq 0$$

where $\mu > 0$.

Note that

$$\begin{aligned}
\sum_{k=0}^{\infty} p_k &= e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} \\
&= e^{-\mu}\ e^{\mu} \\
&= 1
\end{aligned}$$

$\Rightarrow$ the p_k define a set of probabilities!