CH. 3 DISCRETE RANDOM VAR'S AND THEIR PROB DISTRIBUTIONS

Events of interest in an experiment often expressible in terms of a single number (or perhaps a few numbers).

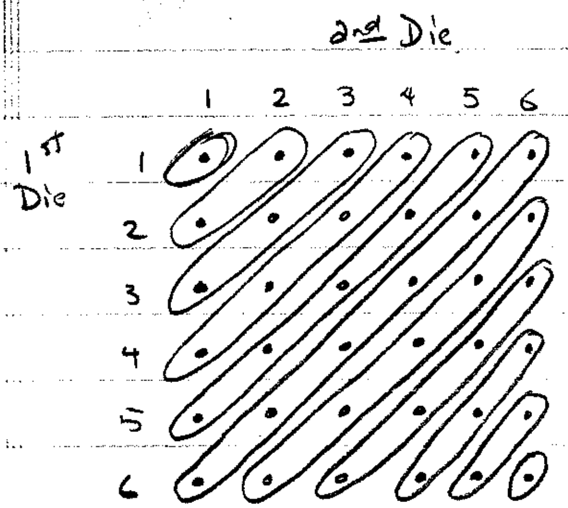
Ex. Toss two coins. $Y = \#$ H's.

Ex. Suppose in rolling two fair dice, we are interested in the sum of the two up faces.

Let $Y =$ sum of two up faces.

Then sample space may be partitioned according to the possible values of Y.

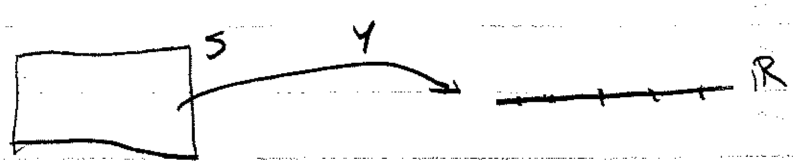
Sample Space



For fair dice:

$P(Y=2) = \frac{1}{36}$	$P(Y=8) = \frac{5}{36}$
$P(Y=3) = \frac{2}{36}$	$P(Y=9) = \frac{4}{36}$
$P(Y=4) = \frac{3}{36}$	$P(Y=10) = \frac{3}{36}$
$P(Y=5) = \frac{4}{36}$	$P(Y=11) = \frac{2}{36}$
$P(Y=6) = \frac{5}{36}$	$P(Y=12) = \frac{1}{36}$
$P(Y=7) = \frac{6}{36}$	

Note that Y is just a real-valued fcn defined over a sample space.



Y is called a <u>random variable</u>. In this case Y is <u>discrete</u>.

A <u>discrete r.v.</u> is one which has only a finite or at most a countably infinite number of possible values.

The set of possible values and their associated probabilities together are called the <u>distribution</u> of the (discrete) random variable Y.
We may write

$$p(y) = P(Y=y).$$

$p(y)$ is then called the <u>probability fcn</u> (p.f.) of Y.

The distn may be specified by a formula, a table, or a graph.

<u>Ex</u>. An urn contains 10 balls, 3 black and 7 white.
3 balls selected at random, without replacement.
Let Y = # of black balls in sample. Find distn of Y.

Y can take values $y = 0, 1, 2, 3$.

$$p(0) = P(Y=0) = \frac{\binom{3}{0}\binom{7}{3}}{\binom{10}{3}} = \frac{35}{120}$$

$$p(1) = P(Y=1) = \frac{\binom{3}{1}\binom{7}{2}}{\binom{10}{3}} = \frac{63}{120}$$

$$p(2) = P(Y=2) = \frac{\binom{3}{2}\binom{7}{1}}{\binom{10}{3}} = \frac{21}{120}$$

$$p(3) = P(Y=3) = \frac{\binom{3}{3}\binom{7}{0}}{\binom{10}{3}} = \frac{1}{120}$$

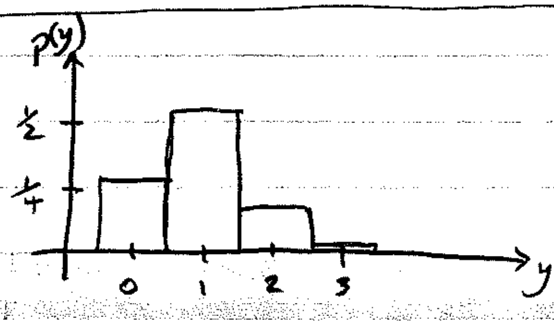
<u>Formula</u>

$$p(y) = \frac{\binom{3}{y}\binom{7}{3-y}}{\binom{10}{3}}, \quad y = 0, 1, 2, 3.$$

<u>Table</u>

y	$p(y)$
0	$35/120$
1	$63/120$
2	$21/120$
3	$1/120$
	$\underline{1}$

<u>Graph</u>



Note: $\sum_y p(y) = \underline{1}$.

EXPECTED VALUES

The expected value (mean) of a discrete r.v. Y with p.f. $p(\cdot)$ is defined to be

$$\mu = E(Y) = \sum_y y \cdot p(y),$$

where the sum is over all possible values of y.

Discuss why this makes sense. Long run average (mean). Center of Mass/Gravity of Dist.

Ex. Rolling two fair dice. Y = sum of up faces.

y	$p(y)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

$$E(Y) = \sum_{y=2}^{12} y \cdot p(y)$$

$$= (2)\left(\frac{1}{36}\right) + (3)\left(\frac{2}{36}\right) + (4)\left(\frac{3}{36}\right) + (5)\left(\frac{4}{36}\right) + (6)\left(\frac{5}{36}\right) + (7)\left(\frac{6}{36}\right) + (8)\left(\frac{5}{36}\right) + (9)\left(\frac{4}{36}\right) + (10)\left(\frac{3}{36}\right) + (11)\left(\frac{2}{36}\right) + (12)\left(\frac{1}{36}\right)$$

$$= \frac{1}{36}[252] = 7.$$

No surprise.

Ex. 10 balls, 3 black, 7 white. Draw 3 at random without replacement.
Y = # black in sample.

$$E(Y) = (0)\left(\frac{35}{120}\right) + (1)\left(\frac{63}{120}\right) + (2)\left(\frac{21}{120}\right) + (3)\left(\frac{1}{120}\right)$$

$$= \frac{1}{120}\left(0 + 63 + 42 + 3\right) = \frac{108}{120} = \frac{9 \times \cancel{12}}{10 \times \cancel{12}} = \frac{9}{10}.$$

Note: "expected" value not necessarily a possible value.

THM. (The Law of Unconscious Statistician, Discrete Version)
Let Y be a discrete r.v with p.f. $p(y)$, and let $g(Y)$ be a real-valued fcn of Y. Then the expected value of $g(Y)$ is

$$E[g(Y)] = \sum_y g(y)p(y). \quad \neq g(E(Y))$$

Note: What is the content of this "theorem"? Explain it. Give a simple example.
$Y = -1, 0, 1$ with prob $= \frac{1}{4}, \frac{1}{2}$.
$X = Y^2 = 0, 1$ with prob $\frac{1}{2}$.

Proof: (Not completely rigorous). Let $g_1, g_2, \ldots$ be the possible values of $g(Y)$ and let

$$A_i = \{ y_j : g(y_j) = g_i \}.$$

Then the p.f. of $g(Y)$ is

$$p^*(g_i) = P(g(Y) = g_i) = \sum_{y_j \in A_i} p(y_j)$$

and

$$E[g(Y)] = \sum_i g_i \cdot p^*(g_i) \qquad (\text{defn})$$

$$= \sum_i \Big(g_i \cdot \sum_{y_j \in A_i} p(y_j) \Big)$$

$$= \sum_i \sum_{y_j \in A_i} g_i \cdot p(y_j) = \sum_i \sum_{y_j \in A_i} g(y_j)\, p(y_j)$$

$$= \sum_y g(y)\, p(y). \quad \square$$

$\ast$ NOTE: $E[g(Y)] \neq g(E(Y))$.

Ex. In the last example, find $E(Y^2)$.

$$E(Y^2) = \sum_y y^2 p(y) = (0^2)\left(\frac{35}{120}\right) + (1^2)\left(\frac{63}{120}\right) + (2^2)\left(\frac{21}{120}\right) + (3^2)\left(\frac{1}{120}\right)$$

$$= \frac{1}{120}(0 + 63 + 84 + 9) = \frac{156}{120} = \frac{13 \cdot 12}{10 \cdot 12} = \frac{13}{10}$$

Note: $E(Y^2) \neq [E(Y)]^2 = \left(\frac{9}{10}\right)^2 = \frac{81}{100}$.

Ex. Find $E\left(\frac{1}{Y+1}\right)$. Note: $E\left(\frac{1}{Y+1}\right) \neq \frac{1}{E(Y)+1} = \frac{1}{\frac{9}{10}+1} = \frac{10}{19}$.

$$E\left(\frac{1}{Y+1}\right) = \sum_y \left(\frac{1}{y+1}\right) p(y)$$

$$= \left(\frac{1}{0+1}\right)\left(\frac{35}{120}\right) + \left(\frac{1}{1+1}\right)\left(\frac{63}{120}\right) + \left(\frac{1}{2+1}\right)\left(\frac{21}{120}\right) + \left(\frac{1}{3+1}\right)\left(\frac{1}{120}\right)$$

$$= \frac{1}{120}\left(35 + \frac{63}{2} + 7 + \frac{1}{4}\right) = \frac{295}{480} = \frac{59}{96}$$

DEFN. The variance of Y is given by

$$\sigma^2 = V(Y) := E\left[(Y-\mu)^2\right] \qquad \left(\text{where } \mu = E(Y)\right)$$

The standard deviation of Y is $\sigma = \sqrt{V(Y)}$. (measure of dispersion)

Ex. Calculate the variance of Y in the last example.

$$\sigma^2 = V(Y) = E\left[(Y-\mu)^2\right] = \sum_{y=0}^{3} (y-\mu)^2 p(y)$$

$$= \left(0 - \frac{9}{10}\right)^2\left(\frac{35}{120}\right) + \left(1 - \frac{9}{10}\right)^2\left(\frac{63}{120}\right) + \left(2 - \frac{9}{10}\right)^2\left(\frac{21}{120}\right) + \left(3 - \frac{9}{10}\right)^2\left(\frac{1}{120}\right)$$

$$= \left(\frac{1}{120}\right)\left(\frac{1}{100}\right)\left[(81)(35) + (1)(63) + (121)(21) + (441)(1)\right] = \frac{5880}{(120)(100)} = \boxed{\frac{49}{100}}$$

-71

OTHER RULES OF EXPECTATIONS

(1) If c is a constant, then $E(c) = c$.
(2) If c is a constant, then $E[c \cdot g(Y)] = c \cdot E[g(Y)]$.

(3) $E[g_1(Y) + \cdots + g_k(Y)] = E[g_1(Y)] + \cdots + E[g_k(Y)]$.

THEOREM. $\sigma^2 = V(Y) = E(Y^2) - \mu^2$.

Proof. $V(Y) = E[(Y-\mu)^2] = E[Y^2 - 2\mu Y + \mu^2]$

$$= E(Y^2) - 2\mu E(Y) + \mu^2$$

$$= E(Y^2) - 2\mu \cdot \mu + \mu^2 = E(Y^2) - \mu^2.$$

Ex. Recall in the last example, $V(Y) = \frac{49}{100}$.
But we also calculated that $E(Y^2) = 13/10$ and $\mu = 9/10$, so by the theorem

$$\sigma^2 = V(Y) = \frac{13}{10} - \left(\frac{9}{10}\right)^2 = \frac{130-81}{100} = \frac{49}{100}.$$

St. dev. is $\sigma = \sqrt{V(Y)} = 7/10$.

3.4 Binomial Distribution

* Binomial Experiment

- n trials (n fixed)
- Each trial results in S or F
- Trials independent
- $p \equiv$ prob. of S in an individual trial, same for each trial
- $Y =$ # S's in the n trials

Possible values of Y: $y = 0, 1, 2, \ldots, n$

Typical sample point: $\underbrace{FSSFFFFSF \cdots FS}_{n \text{ letters}}$

Probability of a particular sample point with y S's and $(n-y)$ F's:

$$p^y q^{n-y}, \quad \text{where } q = 1-p = \text{Prob of F in a single trial.}$$

e.g., if $n=5$,

$$P(\{SFFSF\}) = pqqpq = p^{\overset{y=2}{2}} q^{\overset{n-y=5-2=3}{3}}$$

Why? What assumption(s) are we using here? All of them really.

How many sample points have y S's and $n-y$ F's?

$$\underbrace{\_\ \_\ \_\ \cdots\ \_\ \_}_{n \text{ slots}} \quad \leftarrow \text{Fill in with } y \text{ S's and } n-y \text{ F's.}$$

$\binom{n}{y}$ ways to do this. (Choose y slots to hold the S's)

Prob funct. of Y:

$$p(y) = \binom{n}{y} p^y q^{n-y} \quad , \quad y = 0, 1, \ldots, n$$
$$0 \le p \le 1.$$

write:

$$Y \sim Bin(n, p)$$

Note:

Binomial Theorem

$$\sum_{y=0}^{n} p(y) = \sum_{y=0}^{n} \binom{n}{y} p^y q^{n-y} = (p+q)^n = 1^n = 1$$

Ex: Toss biased coin 8 times, $P(H) = 1/3$.

(i) What is prob to get at least 6 H's?

$Y = \#$ H's in 8 tosses

$\Rightarrow Y \sim Bin(n=8, p=1/3)$.

$$P(Y \ge 6) = p(6) + p(7) + p(8)$$
$$= \binom{8}{6}\left(\tfrac{1}{3}\right)^6\left(\tfrac{2}{3}\right)^2 + \binom{8}{7}\left(\tfrac{1}{3}\right)^7\left(\tfrac{2}{3}\right) + \binom{8}{8}\left(\tfrac{1}{3}\right)^8\left(\tfrac{2}{3}\right)^0$$
$$= \frac{43}{2187} \approx .02$$

(ii) What is prob at least 1 H?

$$P(Y \ge 1) = 1 - P(Y=0) = 1 - \left(\tfrac{2}{3}\right)^8$$

Note: Table I in Appendix gives $P(Y \le y)$.

MEAN AND VARIANCE OF THE BINOMIAL

Thm. If $Y \sim \text{Bin}(n,p)$, then

$$\underbrace{E(Y) = np}_{\text{What else could it be?}} \qquad \text{and} \qquad V(Y) = npq.$$

Proof.
$$E(Y) = \sum_y y\,p(y) = \sum_{y=0}^{n} y \binom{n}{y} p^y q^{n-y}$$
$$= \sum_{y=1}^{n} y \binom{n}{y} p^y q^{n-y}$$
$$= \sum_{y=1}^{n} y \cdot \frac{n!}{y!(n-y)!} p^y q^{n-y}$$
$$= \sum_{y=1}^{n} \frac{n!}{(y-1)!(n-y)!} p^y q^{n-y}$$
$$= np \cdot \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!\,[(n-1)-(y-1)]!} p^{y-1} q^{(n-1)-(y-1)}$$
$$= np \cdot \sum_{y=1}^{n} \binom{n-1}{y-1} p^{y-1} q^{(n-1)-(y-1)}$$

$x = y-1$
$$= np \cdot \underbrace{\sum_{x=0}^{n-1} \binom{n-1}{x} p^x q^{(n-1)-x}}_{=1 \text{ (sum of all Bin}(n-1,p)\text{ probabilities)}}$$

$$= np.$$

Aside: $E(Y^2) = \sum_{y=0}^{n} y^2 \cdot \binom{n}{y} p^y q^{n-y} = \sum_{y=1}^{n} y^2 \frac{n!}{y!(n-y)!} p^y q^{n-y} = ?$

Note that y^2 is not a factor of $y!$

Consider

$$E[Y(Y-1)] = \sum_y y(y-1)p(y) = \sum_{y=0}^{n} y(y-1)\binom{n}{y}p^y q^{y-1}$$

$$= \sum_{y=2}^{n} y(y-1)\frac{n!}{y!(n-y)!}p^y q^{n-y}$$

$$= \sum_{y=2}^{n} \frac{n!}{(y-2)!(n-y)!}p^y q^{n-y}$$

$$= n(n-1)p^2 \cdot \sum_{y=2}^{n} \frac{(n-2)!}{(y-2)![(n-2)-(y-2)]!}p^{y-2}q^{(n-2)-(y-2)}$$

$$= n(n-1)p^2 \sum_{y=2}^{n} \binom{n-2}{y-2} p^{y-2} q^{(n-2)-(y-2)}$$

$x = y-2$

$$= n(n-1)p^2 \underbrace{\sum_{y=0}^{n-2} \binom{n-2}{x} p^x q^{(n-2)-x}}_{=1 \text{ (sum of all Bin}(n-2,p)\text{ probs.)}}$$

Skip

Similar arguments give:

$$E[Y(Y-1)] = n(n-1)p^2.$$

Now, $E[Y(Y-1)] = E[Y^2 - Y] = E(Y^2) - E(Y)$

$$\Rightarrow E(Y^2) = E[Y(Y-1)] + E(Y) = n(n-1)p^2 + np.$$

Thus,

$$V(Y) = E(Y^2) - [E(Y)]^2 = n(n-1)p^2 + np - (np)^2$$

$$= n^2p^2 - np^2 + np - n^2p^2 = np - np^2$$

$$= np(1-p) = npq. \quad \blacksquare$$

Example. In ~~the triangle taste test~~ unfair coin example,

$$Y \sim \text{Bin}(n=8,\ p=1/3)$$

$$\mu = E(Y) = np = 8/3$$

$$\sigma^2 = V(Y) = npq = (8)(1/3)(2/3) = 16/9$$

$$\sigma = 4/3.$$

3.5 GEOMETRIC DIST $\left(Y \sim \text{Geom}(p) \right)$.

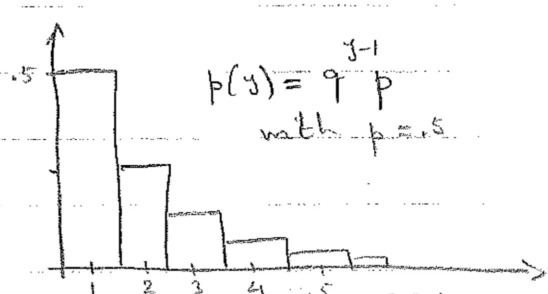
Exp. is similar to binomial: Independent, identical trials, each resulting in S with prob p and F with prob $q = 1-p$, $0 < p \le 1$.

Often used to model distribution of waiting times until an event happens.

Y = # of trials required to obtain first success.

S	$Y=1$	$p(1) = p$
FS	$Y=2$	$p(2) = qp$
FFS	$Y=3$	$p(3) = q^2 p$
$\vdots$	$\vdots$	$\vdots$
$\underbrace{FF\cdots F}_{y-1}\underbrace{S}_{1}$	$Y=y$	$p(y) = q^{y-1}p, \quad y = 1,2,\ldots$

Note:
$$\sum_{y=1}^{\infty} p(y) = \sum_{y=1}^{\infty} q^{y-1}p$$
$$= p\sum_{k=0}^{\infty} q^k \qquad (k = y-1)$$
$$= p \cdot \frac{1}{1-q} = \underline{1}.$$



See graph on p. 100 for $p=\frac{1}{2}$!

or roll a die until 1st 6 occurs

Ex. Russian roulette with a 6-shot revolver and 1 bullet. What is the distribution of the number of trigger pulls needed to fire the bullet?

$p = \frac{1}{6}$, $q = \frac{5}{6}$ $\qquad p(y) = \left(\frac{5}{6}\right)^{y-1}\left(\frac{1}{6}\right)$, $y = 1,2,3,\ldots$

What is the prob that the gun will not have fired in the first 3 trigger pulls? (ctd)

Ex Russian Roulette (ctd)

Easy way: $P(FFF) = \left(\frac{5}{6}\right)^3 = \frac{125}{216}$

or

Harder way:
$$
\begin{aligned}
P(Y>3) &= 1 - P(Y \le 3) \\
&= 1 - \left[p(1) + p(2) + p(3) \right] \\
&= 1 - \left[\tfrac{1}{6} + \left(\tfrac{5}{6}\right)\left(\tfrac{1}{6}\right) + \left(\tfrac{5}{6}\right)^2\left(\tfrac{1}{6}\right) \right] \\
&= \frac{125}{216}
\end{aligned}
$$

or

$$
\begin{aligned}
P(Y>3) = P(Y \ge 4) &= \sum_{y=4}^{\infty} \left(\tfrac{5}{6}\right)^{y-1} \left(\tfrac{1}{6}\right) \\
&= \left(\tfrac{5}{6}\right)^3 \left(\tfrac{1}{6}\right) \sum_{y=4}^{\infty} \left(\tfrac{5}{6}\right)^{y-4} \\
&= \left(\tfrac{5}{6}\right)^3 \left(\tfrac{1}{6}\right) \frac{1}{1 - 5/6} = \left(\tfrac{5}{6}\right)^3 = \frac{125}{216}.
\end{aligned}
$$

What is the expected number of tries to get the 1^{st} shot, $E(Y)$?

Thm. For $Y \sim Geom(p)$, $E(Y) = \frac{1}{p}$ and $V(Y) = \frac{q}{p^2}$.

Proof.

One way: $E(Y) = \sum_y y p(y) = \sum_{y=1}^{\infty} y \cdot q^{y-1} p = p \underbrace{\sum_{y=1}^{\infty} y \cdot q^{y-1}}_{\text{? how to get this?}}$

Now note that
$$\frac{d}{dq}\left(\sum_{y=0}^{\infty} q^y \right) = \frac{d}{dq}\left(\frac{1}{1-q} \right)$$

and
$$\text{LHS} = \sum_{y=0}^{\infty} \frac{d}{dq}(q^y) = \sum_{y=1}^{\infty} y q^{y-1} \qquad \left(\frac{d}{dq}(q^0) = \frac{d}{dq}(1) = 0 \right)$$

while
$$\text{RHS} = \frac{d}{dq}\left[(1-q)^{-1}\right] = (-1)(1-q)^{-2}(-1) = \frac{1}{(1-q)^2}$$

(ctd)

Proof (ctd) Thus $\sum_{y=1}^{\infty} y \cdot q^{y-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$.

and

$$E(Y) = p \cdot \sum_{y=1}^{\infty} y \cdot q^{y-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

ANOTHER WAY: See the "perturbation method" in Concrete Mathematics, Graham, Knuth, and Patashnik.

For $V(Y)$, we use a similar trick as in the binomial case and look at $E[Y(Y-1)]$, and use 2^{nd} derivatives this time:

$$E[Y(Y-1)] = \sum_{y=1}^{\infty} y(y-1)q^{y-1}p = pq \sum_{y=2}^{\infty} y(y-1)q^{y-2}$$

Continue by considering

$$\frac{d^2}{dq^2}\left(\frac{1}{1-q}\right) = \frac{d^2}{dq^2}\left(\sum_{y=0}^{\infty} q^y\right) = \sum_{y=2}^{\infty} y(y-1)q^{y-2} = \frac{2}{(1-q)^3}.$$

The rest is homework !!! (Skip...)

$$\frac{d^2}{dq^2}\left(\frac{1}{1-q}\right) = \frac{d^2}{dq^2}\left[(1-q)^{-1}\right] = \frac{d}{dq}\left[(1-q)^{-2}\right] = \frac{2}{(1-q)^3}$$

$$\Rightarrow E[Y(Y-1)] = pq \cdot \frac{2}{(1-q)^3} = pq \cdot \frac{2}{p^3} = \frac{2q}{p^2}$$

$$\Rightarrow E(Y^2) = E[Y(Y-1)] + E(Y) = \frac{2q}{p^2} + \frac{1}{p}$$

$$\Rightarrow V(Y) = E(Y^2) - \mu^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2}$$

$$= \frac{2q - (1-p)}{p^2} = q/p^2 \quad \square$$

Ex. In the roulette example, with $Y \sim Geom(1/6)$.

$$E(Y) = 6 \quad \text{and} \quad V(Y) = \frac{5/6}{(1/6)^2} = 30.$$

(Skip Summer) (Skip problems 3.54 - 3.58)

---

§ 3.6 NEGATIVE BINOMIAL DISTRIBUTION

Again we consider independent identical trials with fixed prob p of S and $q = 1-p$ of F in each trial. But here we let

$$Y = \text{\# of trials required to obtain the } r^{\text{th}} \text{ success}.$$

$$Y \sim NB(r,p) \quad , \quad 0 < p \le 1.$$

$Y=y$:

$$\overbrace{\underbrace{FSFFSF \cdots SSF}_{y-1 \text{ trials}}S}^{y \text{ trials},\ y \ge r}$$

last trial must be a S

with $\begin{cases} r-1 & S\text{'s} \\ y-r & F\text{'s} \end{cases}$

Each such sequence has prob. r S's and $y-r$ F's, so has prob.

$$q^{y-r} p^r \quad , \quad y \ge r.$$

There are $\binom{y-1}{r-1}$ such sequences (Must have $r-1$ S's scattered among first $y-1$ trials.)

So

$$p(y) = \binom{y-1}{r-1} q^{y-r} p^r, \qquad y = r, r+1, r+2, \ldots$$

NB: $NB(1,p) = Geom(p)$.

Extra
(58.5)

<u>CRAPS</u> (Skip)

Repeat an experiment over and over. Each time get either
A, B, or C with probs. $p_1, p_2, p_3 = 1 - p_1 - p_2$.
Continue till observe either A or B.
What is prob. get A before B?

A	p_1	$= p_3^0 p_1$
CA	$p_3 p_1$	$= p_3^1 p_1$
CCA	$p_3^2 p_1$	$p_3^2 p_1$
CCCA	$p_3^3 p_1$	$p_3^3 p_1$
⋮	⋮	⋮

$$P(A \text{ before } B) = \sum_{k=0}^{\infty} p_3^k p_1 = p_1 \frac{1}{1-p_3} = \frac{p_1}{p_1+p_2}.$$

(c) $$P(4 \text{ before } 7) = \frac{3/36}{\frac{3}{36}+\frac{6}{36}} = \frac{3}{9} = \frac{1}{3}.$$

$F_k = \{\text{roll } k \text{ on first roll}\}, \quad k = 2, \ldots, 12$

$$P(W) = \sum_{k=2}^{12} P(F_k) P(W \mid F_k)$$

3,6 6,3
4,5 5,4

$P(W \mid F_7) = P(W \mid F_{11}) = 1$
$P(W \mid F_2) = P(W \mid F_3) = P(W \mid F_{12}) = 0$
$P(W \mid F_4) = \frac{3}{9} = \frac{1}{3}$ $P(W \mid F_6) = \frac{5}{11}$ $P(W \mid F_9) = \frac{4}{10} = \frac{2}{5}$
$P(W \mid F_5) = \frac{4}{10} = \frac{2}{5}$ $P(W \mid F_8) = \frac{5}{11}$ $P(W \mid F_{10}) = \frac{3}{9} = \frac{1}{3}$

Thm. For $Y \sim NB(r,p)$, $E(Y) = \frac{r}{p}$, $V(Y) = \frac{rq}{p^2}$.

Ex. A smoker has a pack of 20 cigarettes. Normally smokes 1 cig/hr. Trying to cut down.
Scheme: Each hour roll a fair die. If a 1 or a 2 comes up, he will not smoke a cig that hour; otherwise he does.
Let Y = # of hours that the pack lasts.
Then

$$Y \sim NB(20, \tfrac{2}{3})$$

$$\Rightarrow p(y) = \binom{y-1}{r-1} q^{y-r} p^r = \binom{y-1}{19} q^{y-20} p^{20}$$

for $y = 20, 21, 22, \ldots$.

Note that $E(Y) = \frac{r}{p} = 20 \cdot \frac{1}{2/3} = 20 \cdot \frac{3}{2} = 30$

and $V(Y) = \frac{rq}{p^2} = 20 \frac{1/3}{(2/3)^2} = 20 \cdot \frac{1}{3} \cdot \frac{9}{4} = 15$.

what is the prob that the pack lasts at least 24 hours?

$$P(Y \geq 24) = 1 - P(Y < 24) = 1 - [p(20) + p(21) + p(22) + p(23)]$$

$$= 1 - \left[\binom{19}{19}\left(\tfrac{1}{3}\right)^0\left(\tfrac{2}{3}\right)^{20} + \binom{20}{19}\left(\tfrac{1}{3}\right)^1\left(\tfrac{2}{3}\right)^{20} - \binom{21}{19}\left(\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)^{20} + \binom{22}{19}\left(\tfrac{1}{3}\right)^3\left(\tfrac{2}{3}\right)^{20}\right]$$

$$= 1 - \left(\tfrac{2}{3}\right)^{20}\left[1 + \tfrac{20}{3} + \tfrac{210}{9} + \tfrac{1540}{27}\right]$$

$$\approx 1 - .0265 = \underline{.9735}$$

Ex. John wins about $\frac{1}{3}$ of the points when he plays table tennis against Kate. The first player to reach 21 pts wins. Suppose the points played are independent of each other. What is the probability that John will win his next game against Kate?

[NOTE: Not requiring win by two!]

Let

$$Y = \text{\# of points required for John to get 21 points} \sim NB(21, \tfrac{1}{3})$$

Then

$$\{\text{John wins game}\} = \{Y \le 41\}$$

21 pts → John
20 " → Kate

$$\Rightarrow P(\text{John wins}) = P(Y \le 41)$$

$$= \sum_{y=21}^{41} p(y) = \sum_{y=21}^{41} \binom{y-1}{20} \left(\tfrac{2}{3}\right)^{y-21} \left(\tfrac{1}{3}\right)^{21}$$

$$\approx \underline{.0135}$$

[Normal approx gives .0279]

~~ANOTHER WAY:~~

Yes, do this →

Let X = # of points John wins in 1st 41. Then $X \sim Bin(41, \frac{1}{3})$ and John wins if $X \ge 21$.

$$P(X \ge 21) = \cdots = \underline{.0135}$$

[Normal approx gives .0118]

→ Maybe do this with the first player to get 5 points winning? Then

Exercise

$$Y \sim NB(5, \tfrac{1}{3})$$

and

$$P(J \text{ wins}) = P(Y \le 9) = \sum_{y=5}^{9} p(y) = \underline{.1448}$$

OR

$$X \sim Bin(9, \tfrac{1}{3})$$

and $P(J \text{ wins}) = P(X \ge 5) = \cdots = .1448$

Sampling from a finite pop. of S's & F's

Suppose we have N balls, r red, $N-r$ blue.
Plan to sample n, and let Y = no. red in sample.

Sample with replacement: $Y \sim \text{Bin}(n, p)$, $p = \frac{r}{N}$.

Sample without replacement: $Y \sim$ Hypergeometric (later)
(But HG $\to$ Bin as $N \to \infty$) (

Ex. $N = 10$, $r = 4$, $n = 2$.
$R_i = \{ i^{th}$ ball drawn is red $\}$ $\leftarrow$ Success on i^{th} trial.
Sampling without replacement.
$P(R_1) = \frac{4}{10}$
$P(R_2|R_1) = \frac{3}{9}$
Law of Total Prob $\to$ $P(R_2) = P(R_1)P(R_2|R_1) + P(\bar{R}_1)P(R_2|\bar{R}_1)$
$= \left(\frac{4}{10}\right)\left(\frac{3}{9}\right) + \left(\frac{6}{10}\right)\left(\frac{4}{9}\right) = \frac{4}{10}\left(\frac{3}{9} + \frac{6}{9}\right) = \frac{4}{10}$

So prob of S is same for each trial, but trials are dependent.
(Because: $P(R_2) = \frac{4}{10} \neq \frac{3}{9} = P(R_2|R_1)$ & $P(R_2) \neq \frac{4}{9} = P(R_2|\bar{R}_1)$)

Exper: Sample n w/o replacement $p(y) = P(Y = y)$
r = Red, $N-r$ = Blue, Y = # R in sample.

	y	$N=10, r=4, n=2$	$N=100, r=40, n=2$	$N=1000, r=400, n=2$		$\text{Bin}(n=2, p=.4)$
2B :	0	$\left(\frac{6}{10}\right)\left(\frac{5}{9}\right) = \frac{1}{3} = .3\overline{3}$	$.35\overline{75}$	$.359\overline{759}$	$\to$	$.36$
1R, 1B :	1	$\left(\frac{6}{10}\right)\left(\frac{4}{9}\right) + \left(\frac{4}{10}\right)\left(\frac{6}{9}\right) = \frac{8}{15} = .53\overline{3}$	$.45\overline{45}$	$.480\overline{480}$	$\to$	$.48$
2R	2	$\left(\frac{4}{10}\right)\left(\frac{3}{9}\right) = \frac{2}{15} = .1\overline{33}$	$.15\overline{75}$	$.159\overline{759}$	$\to$	$.16$

Note that prob. of red at each trial is $\frac{4}{10} = \frac{40}{100} = \frac{400}{1000} = .4$ in each case.

<u>MORAL</u>: When sampling without replacement from a <u>large</u> population of S's and F's, if the sample size n is small relative to the population size N, the binomial dist. can be used as an approximation. Otherwise must use hypergeometric dist (~~Sec 3.8~~ next).

3.7 HYPERGEOMETRIC DIST

Urn contains N balls, r red, $N-r$ white.
Select n balls at random, <u>without</u> replacement.

$Y =$ # red balls in sample. $\quad Y \sim HG(N, r, n)$ (?)

Population	Sample
Size $= N$	Size $= n$
r of type S.	Y of type S.
$N-r$ of type F.	$n-Y$ of type F.

The p.f. of Y is

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}, \qquad \begin{cases} y = 0, 1, \ldots, n & \text{, subject to restrictions:} \\ y \le r \\ n-y \le N-r \end{cases}$$

$$\Rightarrow \max(0, n-(N-r)) \le y \le \min(n, r)$$

<u>Ex.</u> A lot of 24 components arrives at a factory.
Acceptance sampling plan: Select 4 at random for testing.
Reject the lot if any of the 4 is defective.
If the lot actually contains 3 defectives, what is the prob. that it will be rejected?

Let $Y =$ # defectives in sample. $\quad \left.\begin{array}{l} N = 24 \\ r = 3 \\ n = 4 \end{array}\right\} Y \sim HG$

Then
$$p(y) = \frac{\binom{3}{y}\binom{21}{4-y}}{\binom{24}{4}}, \qquad y = 0, 1, 2, 3$$

and
$$P(\text{reject lot}) = P(Y \ge 1) = 1 - p(0) = 1 - \frac{\binom{3}{0}\binom{21}{4}}{\binom{24}{4}} = 1 - \frac{15 \cdot 19}{23 \cdot 22}$$
$$= 1 - .5632411 = \underline{.4368}$$

Note that using binomial here, get $1 - \left(\frac{21}{24}\right)^4 \approx .4138$

Note that $\frac{N-n}{N-1} = 1$ if $r=1$

and $\frac{N-n}{N-1} \to 1$ as $N \to \infty$ if n fixed.

Thm. If $Y \sim HG(N, r, n)$, then

$$E(Y) = \frac{nr}{N} \quad \text{and} \quad V(Y) = \underset{\substack{\text{Sample}\\ \text{size}}}{n} \underset{\substack{\text{Prop.}\\ \text{of S's}\\ \text{in pop.}}}{\left(\frac{r}{N}\right)} \underset{\substack{\text{Prop.}\\ \text{of F's}\\ \text{in pop.}}}{\left(\frac{N-r}{N}\right)} \underset{\substack{\text{Finite pop.}\\ \text{correction}\\ \text{factor.}}}{\left(\frac{N-n}{N-1}\right)}$$

Show what happens as $N \to \infty$ and $\frac{r}{N} \to p$.

Note that this is like

$$V(Y) = \underbrace{npq}_{\substack{\text{binomial}\\ \text{variance}}} \underbrace{\left(\frac{N-n}{N-1}\right)}_{\substack{\text{Finite pop.}\\ \text{correction}}}, \quad p = \frac{r}{N}, \quad q = 1 - \frac{r}{N} = \frac{N-r}{N}.$$

(adjustment when n large relative to N).
f.p.c. $\to 1$ as $N \to \infty$

Ex. In the last example, the mean and variance of the number of defectives in the sample are $(N=24, r=3, n=4)$

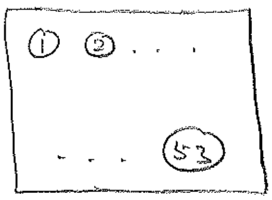
$$E(Y) = n\frac{r}{N} = \frac{(4)(3)}{24} = \frac{1}{2}$$

$$V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = (4)\left(\frac{3}{24}\right)\left(\frac{21}{24}\right)\left(\frac{24-4}{24-1}\right)$$

$$= (4)\left(\frac{1}{8}\right)\left(\frac{7}{8}\right)\left(\frac{20}{23}\right) = \frac{35}{92} \approx \underline{\underline{.3804}}$$

Ex: (~~FL~~ lotto)

Buy ticket with 6 numbers from 1 to 53. Prizes if you match 3, 4, 5, 6 of the weekly randomly drawn #'s.



- $N = 53$
- 6 of these actually match the #'s on your ticket (think of these as red $\Rightarrow r = 6$)
- 6 #'s are randomly drawn from urn w/o replacement $\Rightarrow n = 6$.

Y = # balls matching numbers on your ticket.

$Y \sim HG(N=53,\ r=6,\ n=6)$

$$\mathbb{E}(Y) = \frac{nr}{N} = \frac{6^2}{53} \simeq 0.68$$

$$V(Y) = 6\left(\frac{6}{53}\right)\left(\frac{47}{53}\right)\left(\frac{47}{52}\right) \simeq 0.54$$

y	$p(y) = \binom{r}{y}\binom{N-r}{n-y} \Big/ \binom{N}{n}$			
0	$\binom{6}{0}\binom{47}{6} \Big/ \binom{53}{6}$	$\simeq .468$		0.9993
1	$\binom{6}{1}\binom{47}{5}$	$\simeq .401$		
2	$\binom{6}{2}\binom{47}{4}$	$\simeq .117$		
3	$\binom{6}{3}\binom{47}{3}$	$\simeq .0141$		
4	$\binom{6}{4}\binom{47}{2}$	$\simeq .000706$	$= 7.06 \times 10^{-4}$	
5	$\binom{6}{5}\binom{47}{1}$	$\simeq .0000123$	$= 1.23 \times 10^{-5}$	
6	$\underbrace{\binom{6}{6}\binom{47}{0}}_{1} \Big/ \underbrace{\binom{53}{6}}_{22,957,480}$	$\simeq .0000000436$	$= 4.36 \times 10^{-8}$	

{ A chance of 1 in 23 million to get all 6 right! }

3.8 Poisson Dist

DEFN. A r.v. Y has a Poisson dist with parameter $\lambda > 0$ if Y has p.f.

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda}, \qquad y = 0, 1, 2, \ldots .$$

write

$$Y \sim Poi(\lambda).$$

~~NOTE: Recall that for a function f with derivatives of all orders at a point a, the Taylor series expansion of f about a is~~

$$\cancel{f(x) = f(a) + f'(a)(x-a) + \frac{1}{2} f''(a)(x-a)^2 + \cdots}$$
$$\cancel{= \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(a)(x-a)^k}$$

~~For $a=0$, this is~~

$$\cancel{f(x) = f(0) + f'(0)x + \frac{1}{2} f''(0) x^2 + \frac{1}{6} f'''(0) x^3 + \cdots}$$
$$\cancel{= \sum_{k=0}^{\infty} \frac{1}{k!} f^{(k)}(0) \cdot x^k .}$$

Recall

For the exponential function, $f(x) = \exp(x) = e^x$, we have

$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k .$$

Thus for the Poisson p.f.,

$$\sum p(y) = \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} e^{-\lambda} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{1}{y!} \lambda^y = e^{-\lambda} \cdot e^{\lambda} = \underline{1}.$$

(over) (p. 112)

NOTE: The book shows analytically how the Poisson distribution can be thought of as an approximation to the binomial when n is large and p is small (and $\lambda = np$).

In general this leads us to consider the Poisson dist. when Y counts occurrences of "S", where there are many opportunities for a S to occur, but the probability of S in any given opportunity is small.

Time oriented examples

Examples. ⓪ # of requests to a (or web) database server in a given period.

① # of calls to a switchboard in a given period of time.

② # of automobiles to pass through an intersection in a given time period.

③ # of customers to arrive at a service desk in a given time period.

④ # of particles emitted by a radioactive substance in a given period of time.

⑤ # of firings of a synapse in a given time period.

⑥ # airline accidents in a given time period.

In ⓪–⑥, λ is just the mean number of occurrences in the time period, as we shall see.

Spatial data.

⑦ # of trees of a particular type in a region of forest.

⑧ # of ~~V1 rocket hits in a 1/8 square mile region of~~ typos in ~~London in WWII. ("Gravity's Rainbow" by Thomas Pynchon)~~ your book.

⑨ # of defects in a long length of manufactured cable.

⑩ # of knots in a plank

⑪ # of flaws in a fabric sample

Skip!

Let $n \to \infty$, $p \to 0$ in such a way that $np = \lambda$, i.e., $p = \lambda/n$

If $Y \sim Bin(n,p)$, then for fixed y,

$$p(y) = \binom{n}{y} p^y q^{n-y} = \frac{n!}{y!(n-y)!} \frac{\lambda^y}{n^y} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-y}$$

$$= \frac{\lambda^y}{y!} \frac{n(n-1)(n-2)\cdots(n-y+1)}{n^y} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-y}$$

$$= \frac{\lambda^y}{y!} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{(1)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{y-1}{n}\right)}_{\to 1} \cdot \underbrace{\left(1-\frac{\lambda}{n}\right)^{-y}}_{\to 1}$$

$$\longrightarrow \frac{\lambda^y}{y!} e^{-\lambda} \quad \text{as } n \to \infty.$$

Note also that

$E(Y) = np = \lambda$

$V(Y) = npq = n\left(\frac{\lambda}{n}\right)\left(1-\frac{\lambda}{n}\right) = \lambda\left(1-\frac{\lambda}{n}\right) \longrightarrow \lambda$

THM. If $Y \sim Poi(\lambda)$, then $E(Y) = \lambda$ and $V(Y) = \lambda$.

Pf.

$$E(Y) = \sum_y y \cdot p(y) = \sum_{y=0}^{\infty} y \cdot \frac{\lambda^y}{y!} e^{-\lambda}$$

$$= \sum_{y=1}^{\infty} \frac{\lambda^y}{(y-1)!} e^{-\lambda} = \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1}}{(y-1)!} e^{-\lambda}$$

$$= \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \cdot 1 = \lambda.$$

$$E[Y(Y-1)] = \sum_{y=0}^{\infty} y(y-1) \frac{\lambda^y}{y!} e^{-\lambda}$$

$$= \sum_{y=2}^{\infty} \frac{\lambda^y}{(y-2)!} e^{-\lambda} = \lambda^2 \sum_{y=2}^{\infty} \frac{\lambda^{y-2}}{(y-2)!} e^{-\lambda}$$

$$= \lambda^2 \cdot 1 = \lambda^2$$

$$\Rightarrow E(Y^2) = E[Y(Y-1)] + E(Y) = \lambda^2 + \lambda$$

$$\Rightarrow V(Y) = E(Y^2) - [E(Y)]^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda. \qquad \square$$

Ex. Calls arrive at a switchboard at a rate of 24 calls/hr (on avg).
Assume # of calls received in any given time period is Poisson dist'd.
Find the prob of more than 3 calls in a given 5 minute period.

$$\lambda = \left(\frac{24 \text{ calls}}{\text{hr}}\right)\left(\frac{1}{12}\text{hr}\right) = 2 \text{ calls}/5\text{ min (on avg)}$$

Y = # of calls in 5 minute period : $Y \sim Poi(\lambda = 2)$

$$p(y) = \frac{2^y}{y!}e^{-2}, \quad y = 0, 1, 2, \ldots$$

$$P(Y>3) = 1 - P(Y \le 3) = 1 - [p(0)+p(1)+p(2)+p(3)]$$

$$= 1 - \left[\frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} + \frac{2^2}{2!}e^{-2} + \frac{2^3}{3!}e^{-2}\right]$$

$$= 1 - e^{-2}\left[1+2+2+\frac{4}{3}\right] = 1 - \frac{19}{3}e^{-2}$$

$$\approx 1 - .8571 = .1429$$

OR See Table 3, $[P(Y>3) = 1 - P(Y \le 3) = 1 - .857 = .143]$

What are the mean and variance of Y? $E(Y) = 2 = V(Y)$

Ex. Suppose that the # of choc. chips in a chips ahoy cookie follows a Poisson dist. What should be the average # of chips / cookie in order that not more than one cookie in a thousand should have no choc. chips?

λ = avg # of chips / cookie

Y = # of chips in a randomly selected cookie.

$Y \sim Poi(\lambda)$ and we want to ~~set~~ find smallest λ so that $P(Y=0) \leq .001$

$$P(Y=0) = e^{-\lambda} \leq .001$$

$$\Rightarrow \lambda \geq -\ln(.001) \approx 6.908.$$

Ex. (old) Same setup, but now we wish to have not more than 1 cookie in a hundred to have fewer than 5 chips.

Skip!

want

$$P(Y<5) = P(Y \leq 4) = e^{-\lambda}\left[1+\lambda+\tfrac{1}{2}\lambda^2+\tfrac{1}{6}\lambda^3+\tfrac{1}{24}\lambda^4\right] \overset{!}{=} .01$$

So need to solve the following eqn for λ:

$$e^{-\lambda}\left[1+\lambda+\tfrac{1}{2}\lambda^2+\tfrac{1}{2}\lambda^3+\tfrac{1}{24}\lambda^4\right] = .01$$

Could solve using Bisection method or Newton-Raphson method.

→ Or use Table 3 (pp. 725-9) to get an approximate soln.

Should have about 11.5 to 12 chips / cookie on average.

Ex. Suppose that $Y \sim Bin(600, .005)$.
Find the exact value of $P(Y \le 1)$ and then find an approximation using the Poisson dist.

Exact: $P(Y \le 1) = p(0) + p(1) = \binom{600}{0}(.005)^0(.995)^{600}$

$$+ \binom{600}{1}(.005)^1(.995)^{599}$$

$$\approx .04941 + .14899 = \underline{.1984}$$

Approx: $\lambda = np = (600)(.005) = 3$

$Y \sim Poi(3)$

$$P(Y \le 1) = p(0) + p(1) = e^{-3}[1+3] = 4e^{-3} \approx \underline{.1991}$$

Relative error of only 0.35%

§3.9 MOMENTS AND MOMENT GENERATING FCNS

The i^{th} moment of a r.v. Y is $\mu_i' := E(Y^i)$.

The i^{th} central moment of Y, or the i^{th} moment of Y about its mean is

$$\mu_i := E\left[(Y-\mu)^i\right].$$

Note: $\begin{cases} \mu_1' = \mu \\ \mu_2 = \sigma^2 \end{cases}$

The moment generating function, $m(t)$, of a r.v. Y is

$$m(t) := E(e^{tY}) \quad , \quad t \in \mathbb{R}$$

We say that the m.g.f. of Y "exists" if $\exists\ b > 0 \ni$

$$m(t) < \infty \quad \text{for } |t| < b.$$

Note: ① $m(0) = E(e^0) = 1$ for any r.v. Y.

② For a discrete r.v.,

$$m(t) = \sum_y e^{ty} p(y)$$

(This is similar to a discrete version of the Laplace transform of $p(y)$. Recall

$$\hat{f}(t) = \int_0^\infty e^{-ts} f(s)\, ds$$)

Why is m(t) called the m.g.f.?

Book uses Taylor series expansion of m(t).

Another approach: If the m.g.f. "exists", then it can be shown that for $|t|<b$,

$$m^{(k)}(t) \overset{!}{=} \frac{d^k}{dt^k} m(t) = \frac{d^k}{dt^k} E(e^{tY}) \overset{!}{=} E\left[\frac{d^k}{dt^k}(e^{tY})\right] = E\left[Y^k e^{tY}\right]$$

Thus,

$$m'(t) = \frac{d}{dt} E(e^{tY}) = E\left[\frac{d}{dt} e^{tY}\right] = E\left[Ye^{tY}\right]$$

$$\Rightarrow m'(0) = E\left[Ye^0\right] = E(Y) = \mu_1',$$

$$m''(t) = \frac{d^2}{dt^2} E(e^{tY}) = E\left[\frac{d^2}{dt^2} e^{tY}\right] = E\left[Y^2 e^{tY}\right]$$

$$\Rightarrow m''(0) = E\left[Y^2 e^0\right] = E(Y^2) = \mu_2',$$

and in general,

$$m^{(k)}(0) = \mu_k', \text{ the } k^{\text{th}} \text{ moment of } Y.$$

THM. If the m.g.f. of a r.v. Y "exists", then for any positive integer k,

$$m^{(k)}(0) = \mu_k'.$$

THM. [IMPORTANT] If the m.g.f. of a r.v. "exists", then it completely determines the distribution of the r.v.

So if X and Y have the same m.g.f. (and it "exists"), then X and Y have the same distribution. We will use this later!

<u>Ex</u>. Let $Y \sim Poi(\lambda)$. Then Y has m.g.f.

$$m(t) = E\left(e^{tY}\right) = \sum_{y=0}^{\infty} e^{ty} \cdot p(y)$$

$$= \sum_{y=0}^{\infty} e^{ty} \cdot \frac{\lambda^y}{y!} e^{-\lambda}$$

$$= e^{-\lambda} \sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!}$$

$$e^x = \sum_{y=0}^{\infty} \frac{x^y}{y!}$$

$$= e^{-\lambda} \cdot e^{\lambda e^t} \cdot \underbrace{\sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!} e^{-\lambda e^t}}_{\text{Sum of all Poi}(\lambda e^t)\text{ probs}}$$

$$= e^{-\lambda} \cdot e^{\lambda e^t} \cdot \underline{1} = e^{\lambda(e^t - 1)}.$$

Note that

$$m'(t) = \frac{d}{dt}\left[e^{\lambda(e^t-1)}\right] = e^{\lambda(e^t-1)} \frac{d}{dt}\left[\lambda(e^t-1)\right]$$

$$= e^{\lambda(e^t-1)} \cdot \lambda \cdot e^t$$

$$\Rightarrow \mu = E(Y) = \mu_1' = e^{\lambda(e^0-1)} \cdot \lambda \cdot e^0 = e^0 \cdot \lambda \cdot e^0 = \lambda,$$

and

$$m''(t) = \frac{d^2}{dt^2}\left[e^{\lambda(e^t-1)}\right] = \frac{d}{dt}\left[\lambda e^{\lambda(e^t-1)} \cdot e^t\right]$$

$$= \lambda\left[e^{\lambda(e^t-1)} \cdot e^t + e^t \cdot e^{\lambda(e^t-1)} \cdot \lambda e^t\right]$$

$$= \lambda e^{\lambda(e^t-1)} \cdot e^t \left(1 + \lambda e^t\right)$$

(ctd)

Ex (Poi m.g.f. ctd)

$$\Rightarrow E(Y^2) = \mu_2' = m''(0) = \lambda e^{\lambda(e^0-1)} \cdot e^0 (1+\lambda e^0)$$

$$= \lambda(1+\lambda) = \lambda + \lambda^2$$

$$\Rightarrow \sigma^2 = V(Y) = E(Y^2) - [E(Y)]^2 = \mu_2' - (\mu_1')^2$$

$$= \lambda + \lambda^2 - \lambda^2 = \lambda.$$

Ex. $Y \sim Geom(p)$, $0 < p \le 1$.

$$m(t) = E(e^{tY}) = \sum_y e^{ty} \cdot p(y) = \sum_{y=1}^{\infty} e^{ty} \cdot q^{y-1} \cdot p$$

$$= pe^t \sum_{y=1}^{\infty} e^{t(y-1)} q^{y-1} = pe^t \sum_{y=1}^{\infty} (qe^t)^{y-1}$$

$$= pe^t \cdot \frac{1}{1-qe^t} = \frac{pe^t}{1-qe^t},$$ as long as $qe^t < 1$ $\Leftrightarrow e^t < \frac{1}{q} \Leftrightarrow t < -\log(q)$.

$$m'(t) = \frac{d}{dt}\left[\frac{pe^t}{1-qe^t}\right] = \frac{(1-qe^t)pe^t - pe^t(-qe^t)}{(1-qe^t)^2}$$

$$= \frac{pe^t}{(1-qe^t)^2}$$

$$\Rightarrow \mu = E(Y) = \mu_1' = m'(0) = \frac{pe^0}{(1-qe^0)^2} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

$$m''(t) = \frac{(1-qe^t)^{\cancel{2}} pe^t - pe^t[2\cancel{(1-qe^t)}(-qe^t)]}{(1-qe^t)^{\cancel{4}\,3}}$$

$$= \frac{pe^t[1-qe^t+2qe^t]}{(1-qe^t)^3} = \frac{pe^t(1+qe^t)}{(1-qe^t)^3}$$

(ctd)

Ex (Geometric m.g.f. (ctd))

$$\Rightarrow E(Y^2) = \mu_2' = m''(0) = \frac{pe^0(1+qe^0)}{(1-qe^0)^3}$$

$$= \frac{p(1+q)}{(1-q)^3} = \frac{1+q}{p^2}$$

$$\Rightarrow \sigma^2 = V(Y) = \mu_2 = E(Y^2) - [E(Y)]^2 = \mu_2' - (\mu_1')^2$$

$$= \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Exercises include doing this for the Binomial dist.

See Table inside back cover of text for m.g.f.'s of common distributions.

3.11 TCHEBYSHEFF'S THM
(Chebyshev's Thm)

Recall the *Empirical Rule* for bell-shaped distributions (68%, 95%, almost all). (Chap. 1.)

Explain for data, then for theoretical dists. Mention normal dist

Tchebysheff's Thm gives a lower bound for certain probabilities similar in flavor to the Empirical Rule, but Tchebysheff's result holds regardless of the shape of the distribution.

Thm (Tchebysheff) Let Y be a r.v. with finite mean, μ, and variance, σ^2. Then for any positive constant k,

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2},$$

or equivalently,

$$P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Compare with Empirical Rule - 68%, 95%, ~all

For $k=1$, this says nothing!

For $k=2$: $P(|Y-\mu| < 2\sigma) \geq \frac{3}{4} = 75\%$

$k=3$: $P(|Y-\mu| < 3\sigma) \geq \frac{8}{9} = 88.9\%$

Proof. (LATER)

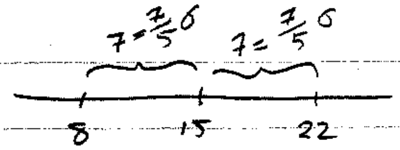
Ex. Suppose we know only that a r.v. Y has mean $\mu = 15$ and variance $\sigma^2 = 25$, and we want to know something about $P(8 < Y < 22)$.

$\frac{8-15}{5} = \frac{-7}{5}$

$\frac{22-15}{5} = \frac{7}{5}$

$\mu = 15 \qquad 8 = 15 - 7 = \mu - \frac{7}{5}\sigma$

$\sigma = 5 \qquad 22 = 15 + 7 = \mu + \frac{7}{5}\sigma$



$$P(8 < Y < 22) = P\left(|Y-\mu| < \tfrac{7}{5}\sigma\right) \geq 1 - \frac{1}{(7/5)^2} = 1 - \frac{25}{49} = \boxed{\frac{24}{49}} = \boxed{.4900}$$

Ex. $Y \sim Poi(\lambda = 64)$. $\mu = \lambda = 64$, $\sigma^2 = \lambda = 64$, $\sigma = 8$

$$P(|Y-64| \geq 10) = P\left(|Y-\mu| \geq \tfrac{10}{8}\sigma\right)$$

$\frac{10}{\sigma} = \frac{10}{8} = \frac{5}{4}$

$$= P\left(|Y-\mu| \geq \tfrac{5}{4}\sigma\right) \leq \frac{1}{(5/4)^2} = \frac{16}{25} = .64$$

$\frac{20}{\sigma} = \frac{20}{8} = \frac{5}{2}$

$$P(|Y-64| \geq 20) = P\left(|Y-\mu| \geq \tfrac{20}{8}\sigma\right)$$

$$= P\left(|Y-\mu| \geq \tfrac{5}{2}\sigma\right) \leq \frac{1}{(5/2)^2} = \frac{4}{25} = .16$$

Note: Exact answers are

$$P(|Y-64| \geq 10) = 1 - P(|Y-64| < 10) = 1 - P(|Y-64| \leq 9)$$

$$= 1 - P(55 \leq Y \leq 73)$$

$$= 1 - [P(Y \leq 73) - P(Y \leq 54)] = 1 - .7655 = .2345$$

$$P(|Y-64| \geq 20) = 1 - [P(Y \leq 83) - P(Y \leq 44)] = 1 - .9853 = .0147$$

Tchebysheff is very conservative for most distributions, but it is a tight inequality in the sense that there always exists a distribution for which equality holds (if $k \geq 1$).

Check this "fact".

Chap. 3 Review

- Dist of r.v. Y characterized by either:
  p.f. $p(y) = P(Y=y)$ or m.g.f. $m(t) = \mathbb{E}\, e^{tY}$.

- Moments:
  $\mu = \mathbb{E}(Y)$, $\sigma^2 = \mathrm{Var}(Y)$, $\mathbb{E}(h(Y)) = \sum_y h(y)p(y)$.

- Cheby's Theom:
  $P(|Y-\mu| < k\sigma) \geq 1 - 1/k^2$ or $P(|Y-\mu| \geq k\sigma) \leq 1/k^2$.

- Recognizing distributions (other than thru p.f. or mgf):

– Y = # trials until r^{th} S: $\to$ Neg. Bin (r, p)
(indep trials, each res in S with prob p) $\to$ Geom $(\cancel{r}\, p)$ = Neg. Bin $(r=1, p)$.

– Y = # trials out of n that result in S:
(~~indep~~ trials, each res in S with prob p)

$\swarrow$ if sampling from ∞ pop of Ss & Fs:
$Y \sim \mathrm{Bin}(n, p)$
(trials indep.)

$\searrow$ if sampling from finite pop of r Ss & $N-r$ Fs:
$Y \sim HG(N, r, n)$
(trials dep. ~~&~~, $p = \frac{r}{N}$)

– Poisson. Counting # events over time/space.
(Approxs Bin(n,p) when n large & p small, $\lambda = np$.)