

3.7 The Hypergeometric Probability Distribution

The *hypergeometric* distribution, the probability of y successes when sampling without¹⁹ replacement n items from a population with r successes and $N - r$ failures, is

$$p(y) = P(Y = y) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}}, \quad 0 \leq y \leq r, \quad 0 \leq n-y \leq N-r,$$

and its expected value (mean), variance and standard deviation are,

$$\mu = E(Y) = \frac{nr}{N}, \quad \sigma^2 = V(Y) = n \left(\frac{r}{N} \right) \left(\frac{N-r}{N} \right) \left(\frac{N-n}{N-1} \right), \quad \sigma = \sqrt{V(Y)}.$$

Exercise 3.7 (The Hypergeometric Probability Distribution)

1. *Hypergeometric: televisions.* Seven television ($n = 7$) tubes are chosen at random from a shipment of $N = 240$ television tubes of which $r = 15$ are defective.

- (a) The probability that $y = 4$ of the chosen televisions are defective is

$$p(4) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}} = (\text{choose one})$$

$$(i) \frac{\binom{15}{4} \times \binom{225}{3}}{\binom{240}{7}} \quad (ii) \frac{\binom{15}{3} \times \binom{225}{4}}{\binom{240}{7}} \quad (iii) \frac{\binom{15}{4} \times \binom{225}{3}}{\binom{240}{7}}$$

- (b) The probability $y = 4$ of the chosen televisions are defective is

$$p(4) = (\text{choose one})$$

- (i) **0.0003069** (ii) **0.0005069** (iii) **0.0006069** (iv) **0.0007069**.

PRGM HPMF ENTER ENTER (again!) 240 ENTER 7 ENTER 15 ENTER 4 ENTER.

- (c) The probability $y = 5$ of the chosen televisions are defective is

$$p(5) = (\text{choose one})$$

- (i) **0.000007069** (ii) **0.000009084** (iii) **0.00010069** (iv) **0.00013059**.

PRGM HPMF ENTER ENTER (again!) 240 ENTER 7 ENTER 15 ENTER 5 ENTER.

¹⁹The hypergeometric assumption of sampling without replacement is more realistic than the binomial assumption of sampling with replacement.

- (d) The probability *at most* $y = 5$ of the chosen televisions are defective is
 $P(Y \leq 5) =$ (choose one)
 (i) **0.900** (ii) **0.925** (iii) **0.950** (iv) **0.999**.
 PRGM HCMF ENTER ENTER (again!) 240 ENTER 7 ENTER 15 ENTER 5 ENTER.
- (e) The probability *at least* $y = 1$ of the chosen televisions are defective is
 $P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) =$ (choose one)
 (i) **0.367** (ii) **0.435** (iii) **0.545** (iv) **0.633**.
 PRGM HCMF ENTER ENTER (again!) 240 ENTER 7 ENTER 15 ENTER 0 ENTER,
 then subtract from 1.
- (f) *Expectation*. The expected number of defective TVs chosen is
 $\mu = E(Y) = \frac{nr}{N} = \frac{(7)(15)}{240} \approx$ (circle one)
 (i) **0.4375** (ii) **0.7375** (iii) **0.9375** (iv) **1.2375**.
- (g) *Expected cost*. If it costs \$75 to repair a defective television, the expected total repair cost is
 $E(C) = E(75Y) = 75E(Y) = 75(0.4375) \approx$ (circle one)
 (i) **\$23.75** (ii) **\$28.75** (iii) **\$32.82** (iv) **\$36.37**.
- (h) *Variance*. The variance in the number of defective TVs chosen
 $\sigma^2 = V(Y) = 7 \left(\frac{15}{240} \right) \left(\frac{240-15}{240} \right) \left(\frac{240-7}{240-1} \right) = n \left(\frac{r}{N} \right) \left(\frac{N-r}{N} \right) \left(\frac{N-n}{N-1} \right) =$
 (i) **0.29986** (ii) **0.39986** (iii) **0.49986** (iv) **0.69986**.
- (i) *Standard deviation in costs*. If it costs \$75 to repair a defective television, the standard deviation in total repair costs is
 $\sigma = \sqrt{V(C)} = \sqrt{V(75Y)} = \sqrt{75^2 V(Y)} = 75\sqrt{V(Y)} \approx 75\sqrt{0.39986} \approx$
 (i) **39.46** (ii) **44.86** (iii) **45.98** (iv) **47.43**.

2. *Binomial approximation to the hypergeometric: televisions*. Seven television ($n = 7$) tubes are chosen at random from a shipment of $N = 240$ television tubes of which $r = 15$ are defective.

- (a) We sample *without* replacement; that is, every time a TV is chosen, we do *not* replace it to be potentially chosen again. In other words, the chance of choosing a defective TV, every time a TV is chosen, *changes* or *depends* on the number of defective TVs that were chosen before it.
 (i) **True** (ii) **False**
- (b) If the sample size, n , is small relative to the number of televisions, N , $\frac{n}{N} < 0.05$, say, the hypergeometric can be approximated by a binomial. The chance, $p = \frac{r}{N}$, of choosing a defective TV, every time a TV is chosen, does not change “that much” when $\frac{n}{N} < 0.05$. Since $\frac{n}{N} = \frac{15}{240} = 0.0625 > 0.05$, the binomial will probably approximate the hypergeometric (choose one)
 (i) **very closely.** (ii) **somewhat closely.** (iii) **not closely at all.**

- (c) Since $p = \frac{r}{N} = \frac{15}{240} = 0.0625$ and so a binomial approximation to:
 $p(5) =$ (i) **0.00157** (ii) **0.00908** (iii) **0.00106** (iv) **0.00139**,

2nd DISTR binompdf 240, 0.0625, 5 ENTER

as compared to 0.000009084 for the hypergeometric,

$$E(Y) = np = (7)(0.0625) =$$

- (i) **0.3998** (ii) **0.4375** (iii) **0.5345** (iv) **0.8345**,

as compared to 0.4375 for the hypergeometric,

$$V(Y) = npq = (7)(0.0625)(1 - 0.0625) \approx$$

- (i) **0.3998** (ii) **0.4102** (iii) **0.7345** (iv) **0.8345**,

as compared to 0.3999 for the hypergeometric.

- (d) In general,

$$\lim_{N \rightarrow \infty} \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} p^y q^{n-y}$$

where $p = \frac{r}{N}$.

- (i) **True** (ii) **False**

3. *Hypergeometric: capture-recapture.* To determine number of perch, N , in Lake Fishalot, $r = 45$ are captured at random from the lake, tagged and let go back into the lake. A short while later, another $n = 32$ perch are captured, of which $y = 2$ are found to be tagged. Approximately how many perch are in Lake Fishalot?

- (a) The chance two of the second group of captured fish are tagged is

$$p(2) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}} = (\text{choose one})$$

$$(i) \frac{\binom{32}{2} \times \binom{N}{43}}{\binom{N}{3200}} \quad (ii) \frac{\binom{45}{2} \times \binom{N-45}{32-2}}{\binom{N}{32}} \quad (iii) \frac{\binom{45}{43} \times \binom{N}{2}}{\binom{N}{3200}},$$

- (b) *Guess* $N = 500$. In this case, chance two of 32 fish chosen are tagged is

$$p(2) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}} = \frac{\binom{45}{2} \times \binom{500-45}{32-2}}{\binom{500}{32}} = (\text{choose one})$$

- (i) **0.24** (ii) **0.26** (iii) **0.29** (iv) **0.32**.

PRGM HPMF ENTER ENTER (again!) 500 ENTER 32 ENTER 45 ENTER 2 ENTER.

- (c) *Guess* $N = 750$. In this case, chance two of 32 fish chosen are tagged is

$$p(2) = \frac{\binom{r}{y} \binom{N-r}{n-y}}{\binom{N}{n}} = \frac{\binom{45}{2} \times \binom{750-45}{32-2}}{\binom{750}{32}} = (\text{choose one})$$

- (i) **0.24** (ii) **0.26** (iii) **0.29** (iv) **0.32**.

PRGM HPMF ENTER ENTER (again!) 750 ENTER 32 ENTER 45 ENTER 2 ENTER.

- (d) *Guess* $N = 1000$. In this case, chance two of 32 fish chosen are tagged is

$$p(2) = (\text{choose one}) \quad (i) \mathbf{0.24} \quad (ii) \mathbf{0.26} \quad (iii) \mathbf{0.29} \quad (iv) \mathbf{0.32}.$$

PRGM HPMF ENTER ENTER (again!) 1000 ENTER 32 ENTER 45 ENTER 2 ENTER.

- (e) A summary of results are given in the following table.

N	500	750	1000
$p(2)$	0.24	0.29	0.26

Since the *largest* chance that 2 of 32 fish chosen are tagged is 0.29, then it seems from the choices given, the number of fish in Lake Fishalot is

$N =$ (choose one) (i) **500** (ii) **750** (iii) **1000** (iv) **1250**.

- (f) The approximation to N would improve if we had more than three $p(2)$ to choose from; however, more effort would be required in calculating the extra $p(2)$. Differentiating

$$\frac{\binom{45}{2} \times \binom{N-45}{32-2}}{\binom{N}{32}}$$

with respect to N and then setting to zero, to locate the maximum N is also possible, but difficult to do.

- (i) **True** (ii) **False**

3.8 The Poisson Probability Distribution

The function of the *Poisson* distribution is

$$p(y) = P(Y = y) = \frac{\lambda^y}{y!} e^{-\lambda}, \quad y = 0, 1, \dots, \lambda > 0,$$

where $e = 2.71828 \dots$ and expected value, variance and standard deviation are,

$$\mu = E(Y) = \lambda, \quad \sigma^2 = V(X) = \lambda, \quad \sigma = \sqrt{\lambda}.$$

Exercise 3.8 (The Poisson Probability Distribution)

1. *Poisson: accidents.* There are an average of $\lambda = 3$ accidents per year along the I-95 stretch of highway between Michigan City, Indiana, and St. Joseph, Michigan.

- (a) The chance there are $y = 4$ accidents is

$$p(4) = \frac{\lambda^y}{y!} e^{-\lambda} = (\text{choose one})$$

$$(i) \frac{3^4}{4!} e^{-3} \quad (ii) \frac{3^4}{4!} e^{-4} \quad (iii) \frac{4^3}{4!} e^{-4} \quad (iv) \frac{4^4}{3!} e^{-3}$$

- (b) The chance there are $y = 4$ accidents is

$$p(4) \approx (\text{choose one}) \quad (i) \mathbf{0.13} \quad (ii) \mathbf{0.17} \quad (iii) \mathbf{0.18} \quad (iv) \mathbf{0.21}.$$

2nd DISTR poissonpdf(ENTER 3, 4) ENTER.

- (c) The chance there are $y = 3$ accidents is

$$p(3) \approx (\text{choose one}) \quad (i) \mathbf{0.17} \quad (ii) \mathbf{0.18} \quad (iii) \mathbf{0.22} \quad (iv) \mathbf{0.24}.$$

2nd DISTR poissonpdf(ENTER 3, 3) ENTER.

- (d) The probability there is *at least* $y = 2$ accidents is

$$P(Y \geq 2) = 1 - P(Y < 2) = 1 - P(Y \leq 1) \approx (\text{choose one})$$

$$p(4) \approx (\text{choose one}) \quad (i) \mathbf{0.78} \quad (ii) \mathbf{0.79} \quad (iii) \mathbf{0.80} \quad (iv) \mathbf{0.81}.$$

Subtract 2nd DISTR poissoncdf(ENTER 3, 1) ENTER from one; notice: poissoncdf, not poissonpdf!

- (e) *Expectation.* The expected number of accidents is

$$\mu = E(Y) = \lambda = (\text{circle one})$$

$$(i) \mathbf{1} \quad (ii) \mathbf{2} \quad (iii) \mathbf{3} \quad (iv) \mathbf{4}.$$

- (f) *Expected cost.* If it costs \$500,000 per accident, the expected yearly cost is $E(C) = E(500000Y) = 500000E(Y) = 500000(3) = (\text{circle one})$

$$(i) \mathbf{\$500,000} \quad (ii) \mathbf{\$1,000,000} \quad (iii) \mathbf{\$1,500,000} \quad (iv) \mathbf{\$2,000,000}.$$

- (g) *Variance.* The variance in the number of accidents per year is

$$\sigma^2 = V(Y) = \lambda = (\text{circle one})$$

$$(i) \mathbf{1} \quad (ii) \mathbf{2} \quad (iii) \mathbf{3} \quad (iv) \mathbf{4}.$$

- (h) *Standard deviation.* Standard deviation in number of accidents per year

$$\sigma = \sqrt{\lambda} = \sqrt{3} \approx (\text{circle one})$$

$$(i) \mathbf{1.01} \quad (ii) \mathbf{1.34} \quad (iii) \mathbf{1.73} \quad (iv) \mathbf{1.96}.$$

2. *Poisson: photons.* A piece of iron is bombarded with electrons and, as a consequence, releases a number of photons. A number, y , of the photon particles released hit a magnetic detection field that surround the piece of iron being tested. It is found that an *average* of $\lambda = 5$ particles hit the magnetic detection field *per microsecond*.

- (a) The chance $y = 2$ particles hit the field per microsecond is

$$p(2) = \frac{\lambda^y}{y!} e^{-\lambda} \approx (\text{choose one}) \quad (i) \mathbf{0.06} \quad (ii) \mathbf{0.07} \quad (iii) \mathbf{0.08} \quad (iv) \mathbf{0.09}.$$

2nd DISTR poissonpdf(ENTER 5, 2) ENTER

- (b) The chance $y = 0$ particles hit the field per microsecond is
 $p(0) \approx$ (choose one) (i) **0.007** (ii) **0.008** (iii) **0.08** (iv) **0.009**.

2nd DISTR poissonpdf(ENTER 5, 0) ENTER

- (c) If an average of $\lambda = 5$ particles hit the field every one microsecond time interval, then, in a *two* microsecond time interval,
 $\lambda = 2 \times 5 =$ (choose one) (i) **10** (ii) **15** (iii) **20** (iv) **25**.

- (d) The chance $y = 3$ particles hit the field in two microseconds is
 $p(3) = \frac{10^3}{3!}e^{-10} \approx$ (choose one)
 (i) **0.007** (ii) **0.008** (iii) **0.009** (iv) **0.010**.

2nd DISTR poissonpdf(ENTER 10, 3) ENTER

- (e) The chance $y = 21$ particles hit the field in four microseconds (so $\lambda = 4 \times 5 = 20$) is $p(21) = \frac{20^{21}}{21!}e^{-20} \approx$ (choose one)
 (i) **0.073** (ii) **0.085** (iii) **0.093** (iv) **0.101**.

2nd DISTR poissonpdf(ENTER 20, 21) ENTER

- (f) *Standard deviation.* Standard deviation in number of field hits per microsecond is $\sigma = \sqrt{\lambda} = \sqrt{5} \approx$ (circle one)
 (i) **1.51** (ii) **1.74** (iii) **2.13** (iv) **2.24**.

3. *Poisson approximation of the binomial: photons.* The Poisson distribution can be used to approximate the binomial distribution by letting $\lambda = np$. This is a fairly good approximation if $np < 7$.

- (a) If $n = 2000$ particles are released by iron per microsecond, and there is a chance $p = 0.005$ that a particle hits the surrounding field per microsecond, then $\lambda = np = 2000(0.005) =$ (circle one)
 (i) **5** (ii) **10** (iii) **15** (iv) **20**.

The chance $y = 15$ particles hit the field in a one microsecond period, is either (*approximate*) *Poisson*²⁰

$$p(15) = \frac{10^{15}}{15!}e^{-10} \approx \text{(circle one)}$$

- (i) **0.00234** (ii) **0.0347** (iii) **0.0445** (iv) **0.0645**

2nd DISTR poissonpdf(ENTER 10, 15) ENTER

or (*exact*) *binomial*

$$p(15) = \frac{2000!}{15!(2000-15)!} \times (0.005)^{15} \times (0.995)^{1985} \approx \text{(circle one)}$$

- (i) **0.00234** (ii) **0.0346** (iii) **0.0445** (iv) **0.0645**.

2nd DISTR binompdf(ENTER 2000, 0.005, 15) ENTER

- (b) If $n = 1000$ and $p = 0.01$, then $\lambda = np = 1000(0.01) =$ (circle one)
 (i) **5** (ii) **10** (iii) **15** (iv) **20**.

Either (*approximate*) *Poisson*

$$p(12) = \frac{10^{12}}{12!}e^{-10} \approx \text{(circle one)}$$

²⁰It is a little surprising Poisson approximation to binomial is close because $\lambda = 2000(0.005) = 10 > 7$.

(i) **0.0948** (ii) **0.1247** (iii) **0.1345** (iv) **0.1445**

2nd DISTR poissonpdf(ENTER 10, 12) ENTER

or (*exact*) *binomial*

$$p(12) = \frac{2000!}{12!(2000-12)!} \times (0.005)^{12} \times (0.995)^{1988} \approx (\text{circle one})$$

(i) **0.0952** (ii) **0.1046** (iii) **0.1245** (iv) **0.1345**.

2nd DISTR binompdf(ENTER 1000, 0.01, 12) ENTER

4. *Poisson process: photons.* A random variable has a Poisson distribution if the Poisson process conditions are satisfied. For the photons example, this would mean the following assumptions are satisfied.

- $P(\text{no "hit" occurs in time subinterval}) = 1 - p$
- $P(\text{one "hit" occurs in time subinterval}) = p$
- $P(\text{two or more "hits" occurs in time subinterval}) = 0$
- Occurrence of hit in each time subinterval is independent of occurrence of event in other nonoverlapping subintervals.

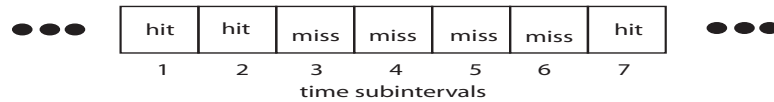


Figure 3.5: Poisson process: sequence of photon hits

- (a) The Poisson process assumes all time subintervals are created so small *one and only one* photon *could* hit the magnetic detection field during each subinterval. During seven time subintervals shown in Figure 3.5, three hits occur in time subintervals 1, 2 and 7 and four misses occur in time subintervals 3, 4, 5 and 6.
- (i) **True** (ii) **False**
- (b) The Poisson process assumes the chance a photon hits the magnetic detection field during any one time subinterval is p and the chance it misses is $q = 1 - p$. It is impossible to have more than one photon hit the magnetic detection field during any time subinterval. Consequently, the time subintervals are *infinitesimally small*.
- (i) **True** (ii) **False**
- (c) The Poisson process assumes a hit in each time subinterval is independent of any other hit in other nonoverlapping subintervals.
- (i) **True** (ii) **False**

- (d) If there was a 20% chance, $p = 0.2$, a photon hit the magnetic detection field during any time subinterval then the chance of $y = 3$ hits in the $n = 7$ time subintervals is given by the *binomial*,

$$p(3) = \binom{n}{y} p^y q^{n-y} = \binom{7}{3} 0.2^3 0.8^{7-3} \approx$$

- (i) **0.089** (ii) **0.115** (iii) **0.124** (iv) **0.134**.

2nd DISTR binompdf(ENTER 7, 0.2, 3) ENTER

- (e) If a large number of time subintervals are considered; in other words, as, n gets bigger, in fact, as $n \rightarrow \infty$, it becomes increasingly difficult to calculate the probability of a number of hits using the binomial. In this case, the poisson is used instead.

- (i) **True** (ii) **False**

- (f) In general,

$$\lim_{n \rightarrow \infty} \binom{n}{y} p^y q^{n-y} = \frac{\lambda^y}{y!} e^{-\lambda}$$

where $\lambda = np$.

- (i) **True** (ii) **False**

3.9 Moments and Moment-Generating Functions

Moment-generating functions, $m(t)$, are useful in calculating the *moments* of the distribution of any²¹ random variable Y . Furthermore, $m(t)$ uniquely identifies any probability distribution. Consequently, it is possible to use either the probability distribution or its associated (and possibly easier to mathematically manipulate) moment-generating function when working with probability distributions.

- The *moment of random variable Y taken about the origin* is defined by,

$$\mu'_k = E(Y^k).$$

- The *moment of random variable Y taken about its mean (or k th central moment of Y)* is defined by,

$$\mu_k = E((Y - \mu)^k).$$

²¹This is true as long as $m(t)$ exists. The function $m(t)$ exists as long as there is a constant b such that $m(t) < \infty$ for $|t| < b$.

- In the discrete case, the moment-generating function of Y is defined by,

$$\begin{aligned}
 m(t) &= E(e^{tY}) = \sum_y e^{ty} p(y) \\
 &= \sum_y \left[1 + ty + \frac{(ty)^2}{2!} + \cdots \right] p(y) \\
 &= \sum_y p(y) + t \sum_y y p(y) + \frac{t^2}{2!} \sum_y y^2 p(y) + \cdots \\
 &= 1 + t\mu'_1 + \frac{t^2}{2!} \mu'_2 + \cdots
 \end{aligned}$$

- Function $m(t)$ “generates” moments of a distribution by successively differentiating $m(t)$ and evaluating the results at $t = 0$,

$$\left. \frac{d^k m(t)}{dt^k} \right|_{t=0} = m^{(k)}(0) = \mu'_k, \quad k = 1, 2, \dots$$

Exercise 3.9 (Moments and Moment-Generating Functions)

1. *Poisson.* Assume moment-generating function for poisson is

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

- (a) Determine $E(Y)$; that is, show $E(Y) = \lambda$ using $m(t)$. First notice

$$E(Y) = E(Y^1) = \mu'_1.$$

So

$$\mu'_1 = m^{(1)}(0) = \left. \frac{d^1 m(t)}{dt^1} \right|_{t=0} = \left. \frac{d e^{\lambda(e^t - 1)}}{dt} \right|_{t=0} = \left[e^{\lambda(e^t - 1)} \cdot \lambda e^t \right]_{t=0} = e^{\lambda(e^0 - 1)} \cdot \lambda e^0 =$$

(i) λ (ii) 2λ (iii) 3λ (iv) $\lambda + 1$.

- (b) Determine $E(Y^2)$. First notice

$$E(Y^2) = \mu'_2.$$

So

$$\mu'_2 = m^{(2)}(0) = \left. \frac{d^2 m(t)}{dt^2} \right|_{t=0} = \left. \frac{d^2 e^{\lambda(e^t - 1)}}{dt^2} \right|_{t=0} = \left[e^{\lambda(e^t - 1)} \cdot (\lambda e^t)^2 + e^{\lambda(e^t - 1)} \cdot \lambda e^t \right]_{t=0}$$

which equals $e^{\lambda(e^0 - 1)} \cdot (\lambda e^0)^2 + e^{\lambda(e^0 - 1)} \cdot \lambda e^0 =$ (choose one)

(i) λ (ii) $\lambda^2 + \lambda$ (iii) $\lambda + 3\lambda$ (iv) $\lambda^3 + \lambda$.

(c) Determine $V(Y)$; that is, show $V(Y) = \lambda$.

$$V(Y) = E(Y^2) - E(Y)^2 = \mu'_2 - (\mu'_1)^2 = (\lambda^2 + \lambda) - \lambda^2 =$$

(i) λ (ii) $\lambda^2 + \lambda$ (iii) $\lambda + 3\lambda$ (iv) $\lambda^3 + \lambda$.

2. *Binomial.* Assume moment-generating function for binomial is

$$m(t) = (pe^t + q)^n.$$

(a) Determine $E(Y)$; that is, show $E(Y) = np$ using $m(t)$.

$$\mu'_1 = m^{(1)}(0) = \left. \frac{d(pe^t + q)^n}{dt} \right|_{t=0} = \left[n(pe^t + q)^{n-1} pe^t \right]_{t=0}$$

which equals $n(pe^0 + q)^{n-1} pe^0 =$ (choose one)

(i) λ (ii) np (iii) $2np$ (iv) npq .

(b) Determine $E(Y^2)$.

$$\mu'_2 = m^{(2)}(0) = \left. \frac{d^2(pe^t + q)^n}{dt^2} \right|_{t=0} = \left[n(n-1)(pe^t + q)^{n-2} (pe^t)^2 + n(pe^t + q)^{n-1} pe^t \right]_{t=0}$$

which is $n(n-1)(pe^0 + q)^{n-2} (pe^0)^2 + n(pe^0 + q)^{n-1} pe^0 =$ (choose one)

(i) $np(n-1)$ (ii) $np^2(n-1)^2 + np$ (iii) $np^2(n-1) + np$.

(c) Determine $V(Y)$; that is, show $V(Y) = npq$.

$$V(Y) = E(Y^2) - E(Y)^2 = \mu'_2 - (\mu'_1)^2 = (np^2(n-1) + np) - (np)^2 = np(1-p) =$$

(i) n (ii) np (iii) $2np$ (iv) npq .

3. *Geometric.* Assume moment-generating function for geometric is

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

(a) Determine $E(Y)$; that is, show $E(Y) = \frac{1}{p}$ using $m(t)$.

$$\mu'_1 = m^{(1)}(0) = \left. \frac{d}{dt} \left(\frac{pe^t}{1 - qe^t} \right) \right|_{t=0} = \left[\frac{pe^t}{(1 - qe^t)^2} \right]_{t=0} =$$

(i) $\frac{1}{p}$ (ii) $\frac{p}{p^3}$ (iii) $\frac{1}{p}$ (iv) $\frac{q}{p}$.

Hint: Since $p + q = 1$, $p = 1 - q$.

(b) Determine $E(Y^2)$.

$$\mu'_2 = m^{(2)}(0) = \left. \frac{d^2}{dt^2} \left(\frac{pe^t}{1 - qe^t} \right) \right|_{t=0} = \left[\frac{(1 - qe^t)^2 pe^t - 2pe^t (1 - qe^t) (-qe^t)}{(1 - qe^t)^4} \right]_{t=0} =$$

(i) $\frac{1}{p}$ (ii) $\frac{p}{p^3}$ (iii) $\frac{1+q}{p^2}$ (iv) $\frac{q-1}{p}$.

(c) Determine $V(Y)$.

$$V(Y) = E(Y^2) - E(Y)^2 = \mu'_2 - (\mu'_1)^2 = \left(\frac{1+q}{p^2}\right) - \left(\frac{1}{p}\right)^2 =$$

$$(i) \frac{q}{p^2} \quad (ii) \frac{p}{p^3} \quad (iii) \frac{1+q}{p^2} \quad (iv) \frac{q-1}{p}.$$

4. Uniqueness of moment-generating functions.

(a) Moment-generating function $m(t) = e^{1(e^t-1)}$ corresponds to (choose one)

(i) **poisson** (ii) **binomial** (iii) **geometric** (iv) **negative binomial**.

Consequently, $\mu = E(Y) = \lambda =$ (circle one)

(i) **1** (ii) **2** (iii) **3** (iv) **4**,

and $\sigma = \sqrt{\lambda} \approx$ (circle one)

(i) **1.00** (ii) **2.41** (iii) **3.41** (iv) **4.41**,

and so $P(0 \leq Y \leq 3) = P(Y \leq 3) \approx$

(circle one) (i) **0.41** (ii) **0.61** (iii) **0.91** (iv) **0.98**.

2nd DISTR poissoncdf(ENTER 1, 3) ENTER

and so the probability Y is within two standard deviations of the mean is,

$$\begin{aligned} P(|Y - \mu| \leq 2\sigma) &= P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) \\ &\approx P(1 - 2(1.41) \leq Y \leq 1 + 2(1.41)) \\ &= P(-1.82 \leq Y \leq 3.82) \\ &= P(0 \leq Y \leq 3) = \end{aligned}$$

(circle one) (i) **0.41** (ii) **0.61** (iii) **0.91** (iv) **0.98**,

2nd DISTR poissoncdf(ENTER 1, 3) ENTER

(b) Moment-generating function $m(t) = (0.2e^t + 0.8)^5$ corresponds to

(i) **poisson** (ii) **binomial** (iii) **geometric** (iv) **negative binomial**.

Consequently, $\mu = E(Y) = np =$ (circle one)

(i) **1** (ii) **2** (iii) **3** (iv) **4**,

and $\sigma = \sqrt{npq} \approx$ (circle one)

(i) **0.41** (ii) **0.51** (iii) **0.89** (iv) **1.21**,

and so the probability Y is within two standard deviations of the mean is,

$$\begin{aligned} P(|Y - \mu| \leq 2\sigma) &= P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) \\ &\approx P(1 - 2(0.89) \leq Y \leq 1 + 2(0.89)) \\ &= P(-0.78 \leq Y \leq 2.78) \\ &= P(0 \leq Y \leq 2) = \end{aligned}$$

(circle one) (i) **0.64** (ii) **0.78** (iii) **0.88** (iv) **0.94**.

2nd DISTR binomcdf(ENTER 5, 0.2, 2) ENTER

- (c) Moment-generating function $m(t) = \frac{0.4e^t}{1-0.6e^t}$ corresponds to
 (i) **poisson** (ii) **binomial** (iii) **geometric** (iv) **negative binomial**.
 Consequently, $\mu = E(Y) = \frac{1}{p} =$ (circle one)
 (i) **1.5** (ii) **2.5** (iii) **3.5** (iv) **4.5**,
 and $\sigma = \sqrt{\frac{q}{p^2}} \approx$ (circle one)
 (i) **1.41** (ii) **1.51** (iii) **1.39** (iv) **1.94**,
 and so the probability Y is within two standard deviations of the mean is,

$$\begin{aligned} P(|Y - \mu| \leq 2\sigma) &= P(\mu - 2\sigma \leq Y \leq \mu + 2\sigma) \\ &\approx P(2.5 - 2(1.94) \leq Y \leq 2.5 + 2(1.94)) \\ &= P(-1.38 \leq Y \leq 6.38) \\ &= P(1 \leq Y \leq 6) = \end{aligned}$$

(circle one) (i) **0.64** (ii) **0.78** (iii) **0.87** (iv) **0.95**.

2nd DISTR geometcdf(ENTER 0.4, 6) ENTER

5. Functions of random variables and associated moment-generating function.

- (a) Moment-generating function for $m(t) = m(0)$.
 For poisson moment-generating function, $m(t) = e^{\lambda(e^t-1)}$,

$$m(0) = e^{\lambda(e^0-1)} =$$

(choose one) (i) **1** (ii) **2** (iii) **3** (iv) **4**.

For geometric moment-generating function, $m(t) = \frac{pe^t}{1-qe^t}$,

$$m(0) = \frac{pe^0}{1-qe^0} =$$

(choose one) (i) **1** (ii) **2** (iii) **3** (iv) **4**.

For negative binomial moment-generating function, $m(t) = \left(\frac{pe^t}{1-qe^t} \right)^r$,

$$m(0) = \left(\frac{pe^0}{1-qe^0} \right)^r =$$

(choose one) (i) **1** (ii) **2** (iii) **3** (iv) **4**.

In general, since $m(t) = E(e^{tY})$,

$$m(0) = E(e^{t(0)}) = E(1) =$$

(choose one) (i) **1** (ii) **2** (iii) **3** (iv) **4**.

(b) *Moment-generating function for $W = 5Y$.*

In general, since $m_Y(t) = E(e^{tY})$,

$$m_W(t) = E(e^{tW}) = E(e^{t(5Y)}) = E(e^{(5t)Y}) =$$

(choose one) (i) $5m_Y(t)$ (ii) $m_Y(5t)$ (iii) $m_W(5t)$ (iv) $m_Y(5t^2)$.

For poisson moment-generating function, $m_Y(t) = e^{\lambda(e^t-1)}$,

$$m_W(t) = m_Y(5t) =$$

(choose one) (i) $e^{\lambda(5e^t-1)}$ (ii) $e^{\lambda(e^{5t}-1)}$ (iii) $e^{\lambda(e^{5t^2}-1)}$ (iv) $e^{\lambda(e^{5t}-1)}$.

For negative moment-generating function, $m_Y(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$,

$$m_W(t) = m_Y(5t) =$$

(i) $\left(\frac{pe^t}{1-qe^t}\right)^r$ (ii) $\left(\frac{5pe^t}{1-qe^{5t}}\right)^r$ (iii) $\left(\frac{pe^{5t}}{1-qe^{5t}}\right)^r$ (iv) $\left(\frac{5pe^{5t}}{1-5qe^{5t}}\right)^r$.

(c) *Moment-generating function for $W = 5Y + 3$.*

In general, since $m_Y(t) = E(e^{tY})$,

$$m_W(t) = E(e^{tW}) = E(e^{t(5Y+3)}) = E(e^{(5t)Y+3t}) = e^{3t} E(e^{(5t)Y}) =$$

(i) $5e^{3t}m_Y(t)$ (ii) $e^{3t}m_Y(5t)$ (iii) $e^{3t}m_W(5t)$ (iv) $e^{3t}m_Y(5t^2)$.

For geometric moment-generating function, $m_Y(t) = \frac{pe^t}{1-qe^t}$,

$$m_W(t) = e^{3t}m_Y(5t) =$$

(i) $e^{3t} \frac{pe^{3t}}{1-qe^t}$ (ii) $e^{5t} \frac{pe^{5t}}{1-qe^t}$ (iii) $e^{3t} \frac{pe^t}{1-5qe^t}$ (iv) $e^{3t} \frac{pe^{5t}}{1-qe^{5t}}$.

For negative moment-generating function, $m_Y(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$,

$$m_W(t) = e^{3t}m_Y(5t) =$$

(i) $\left(\frac{pe^t}{1-qe^t}\right)^r$ (ii) $e^{5t} \left(\frac{5pe^t}{1-qe^{5t}}\right)^r$ (iii) $e^{3t} \left(\frac{pe^{5t}}{1-qe^{5t}}\right)^r$ (iv) $e^{3t} \left(\frac{5pe^{5t}}{1-5qe^{5t}}\right)^r$.

6. *More on moment-generating functions.*

(a) *Deriving binomial moment-generating function.*

$$\begin{aligned} m(t) &= E(e^{tY}) = \sum_{y=0}^n e^{ty} p(y) \\ &= \sum_{y=0}^n e^{ty} \binom{n}{y} p^y q^{n-y} \\ &= \sum_{y=0}^n \binom{n}{y} (pe^t)^y q^{n-y} \\ &= (pe^t + q)^n \end{aligned}$$

(b) *Summary of moment–generating functions.*

DISCRETE	$p(y)$	$m(t)$	μ	σ^2
Binomial	$\binom{n}{y} p^y q^{n-y}$	$(pe^t + q)^n$	np	npq
Poisson	$e^{-\lambda} \frac{\lambda^y}{y!}$	$e^{\lambda(e^t-1)}$	λ	λ
Geometric	$q^{y-1}p$	$\frac{pe^t}{1-qe^t}$	$1/p$	q/p^2
Negative Binomial	$\binom{y-1}{r-1} p^r q^{y-r}$	$\left(\frac{pe^t}{1-qe^t}\right)^r$	r/p	rq/p^2

3.10 Probability–Generating Functions

Not covered.

3.11 Tchebysheff's Theorem

Tchebysheff's Theorem states, for random variable Y with finite μ and σ^2 and for $k > 0$,

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

These two (equivalent) inequalities allow us to specify (very loose) lower bounds on probabilities when the distribution is not known.

Exercise 3.11 (Tchebysheff's Theorem)

1. *Tchebysheff's theorem and binomial: lawyer.* A lawyer estimates she wins 40% ($p = 0.4$) of her cases. Assume each trial is independent of one another and, in general, this problem obeys the conditions of a binomial experiment. The lawyer presently represents 10 ($n = 10$) defendants.

(a) The lawyer's probability of winning is given by the binomial distribution

$$p(y) = \binom{n}{y} p^y q^{n-y}, \quad n = 10, p = 0.4, r = 0, 1, \dots, 10,$$

with corresponding table of

y	0	1	2	3	4	5	6	7	8	9	10
$p(y)$	0.006	0.040	0.121	0.215	0.251	0.201	0.111	0.043	0.011	0.002	0.000

For example, the chance of her winning 4 of 10 cases is (choose one)

(i) **0.121** (ii) **0.215** (iii) **0.251** (iv) **0.351**.

(b) $\mu = np = (10)(0.4) =$ (choose one) (i) **3** (ii) **4** (iii) **5** (iv) **6**.

(c) $\sigma = \sqrt{npq} = \sqrt{(10)(0.4)(0.6)} \approx$ (choose one)

(i) **1.32** (ii) **1.55** (iii) **1.67** (iv) **2.03**.

- (d) According to Tchebysheff's theorem, the probability the number of wins is *within* $k = 2$ standard deviations of the mean number of wins *is at least*

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

- (i) **0.75** (ii) **0.85** (iii) **0.95** (iv) **0.98**.

In actual fact, since $\mu = 4$ and $\sigma \approx 1.55$,

$$\begin{aligned} P(|Y - \mu| < k\sigma) &= P(\mu - k\sigma < Y < \mu + k\sigma) \\ &\approx P(4 - 2(1.55) < Y < 4 + 2(1.5)) \\ &= P(0.9 < Y < 7.1) \\ &= P(1 \leq Y \leq 7) = \end{aligned}$$

- (i) **0.56** (ii) **0.76** (iii) **0.88** (iv) **0.98**.

2nd DISTR binomcdf(ENTER 10, 0.4, 7) ENTER

subtract 2nd DISTR binomcdf(ENTER 10, 0.4, 0) ENTER

Tchebysheff's approximation, 0.75, is a (very) low bound on the actual probability, 0.98.

- (e) According to Tchebysheff's theorem, the probability the number of wins is *within* $k = 2.5$ standard deviations of the mean number of wins *is at least*

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{2.5^2} =$$

- (i) **0.75** (ii) **0.84** (iii) **0.95** (iv) **0.98**.

In actual fact, since $\mu = 4$ and $\sigma \approx 1.55$,

$$\begin{aligned} P(|Y - \mu| < k\sigma) &= P(\mu - k\sigma < Y < \mu + k\sigma) \\ &\approx P(4 - 2.5(1.55) < Y < 4 + 2.5(1.5)) \\ &= P(0.125 < Y < 7.875) \\ &= P(1 \leq Y \leq 7) = \end{aligned}$$

- (i) **0.56** (ii) **0.76** (iii) **0.88** (iv) **0.98**.

Tchebysheff's approximation, 0.84, is a (very) low bound on the actual probability, 0.98.

- (f) According to Tchebysheff's theorem, the probability the number of wins is *beyond* $k = 2.5$ standard deviations from the mean number of wins *is at most*

$$P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2} = \frac{1}{2.5^2} =$$

- (i) **0.12** (ii) **0.16** (iii) **0.25** (iv) **0.34**.

In actual fact, since $\mu = 4$ and $\sigma \approx 1.55$,

$$\begin{aligned} P(|Y - \mu| \geq k\sigma) &= P(Y \leq \mu - k\sigma) + P(Y \geq \mu + k\sigma) \\ &\approx P(Y \leq 4 - 2.5(1.55)) + P(Y \geq 4 + 2.5(1.55)) \\ &= P(Y \leq 0.125) + P(Y \geq 7.875) \\ &= P(Y = 0) + P(8 \leq Y \leq 10) \\ &= 1 - P(1 \leq Y \leq 7) = \end{aligned}$$

(i) **0.01** (ii) **0.02** (iii) **0.03** (iv) **0.04**.

Tchebysheff's approximation, 0.16, is a (very) high bound on the actual probability, 0.02.

2. *Tchebysheff's theorem: Ph levels in soil.* Assume the Ph levels in soil samples taken at Sand Dunes Park, Indiana, have a mean and standard deviation of $\mu = 10$ and $\sigma = 3$ respectively.

- (a) According to Tchebysheff's theorem, the probability the Ph level is *within* $k = 2$ standard deviations of the mean Ph level *is at least*

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

(i) **0.75** (ii) **0.85** (iii) **0.95** (iv) **0.98**.

The actual probability cannot be calculated here because the probability distribution of the Ph levels is unknown in this case. Tchebysheff's approximation, 0.75, will be a low bound on whatever is the actual probability.

- (b) If 400 soil samples are taken, what number will be *at least within* $k = 2$ standard deviations of the mean Ph level? Since $P(|Y - \mu| < 2\sigma) \geq 0.75$, at least $0.75(400) =$ (choose one) (i) **200** (ii) **300** (iii) **400** (iv) **500**.

- (c) According to Tchebysheff's theorem, the probability the Ph level is *beyond* $k = 2.5$ standard deviations from the mean Ph level *is at most*

$$P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2} = \frac{1}{2.5^2} =$$

(i) **0.12** (ii) **0.16** (iii) **0.25** (iv) **0.34**.

Tchebysheff's approximation, 0.16, will be a (very) high bound on whatever is the actual probability.

- (d) According to Tchebysheff's theorem,

$$\begin{aligned} P(5 < Y < 15) &= P(\mu - k\sigma < Y < \mu + k\sigma) \\ &= P\left(10 - \frac{5}{3}(3) < Y < 10 + \frac{5}{3}(3)\right) \\ &\geq 1 - \frac{1}{k^2} = 1 - \frac{1}{\left(\frac{5}{3}\right)^2} = \end{aligned}$$

(circle one) (i) **0.54** (ii) **0.64** (iii) **0.75** (iv) **0.84**.

- (e) If $P(|Y - \mu| \geq k\sigma) \leq 0.35$, then $\frac{1}{k^2} = 0.35$ or $k = \sqrt{\frac{1}{0.35}} \approx$ (choose one)
(i) **1.12** (ii) **1.16** (iii) **1.25** (iv) **1.69**.

3.12 Summary