# Chapter 3 Class Notes

## Sections 3.1 and 3.2

- **Recall, a random variable (RV) is a real-valued function defined over a sample space, S. A random variable Y is said to be discrete if it can take on only a *finite* or *countably infinite* number of distinct values.**
- **Uppercase Y is a random variable, and lowercase y is a particular value that the random variable may assume: Y is random, y is not. (Y = y) is to be understood as the set of points in S assigned to the value y by the RV Y.**
- **Def. 3.2. The probability that Y takes on the value y, P(Y = y), is defined as the *sum of the probabilities of all sample points in S that are assigned to the value y*. Often (usually), we write P(Y = y) as p(y).**
- **Def. 3.3 (p.88). The probability distribution for a discrete RV Y can be represented by a formula, a table, or a graph that provides p(y) = P(Y = y) for all y.**
- **Example 3.1 (p.88). From A = {$M_1$, $M_2$, $M_3$, $W_1$, $W_2$, $W_3$}, we choose just two. There are $\binom{6}{2}$ = 15 sample points in S; Y is the number of women selected. Here the probability distribution formula is $p(y) = \frac{\binom{3}{y}\binom{3}{2-y}}{\binom{6}{2}}$**

for y = 0, 1, 2. For example, p(0) = 3/15 since there are 3 sample points in S that map to Y = 0, viz, $M_1M_2$, $M_1M_3$, $M_2M_3$. The table and histogram are on p.89.

- **Theorem 3.1** (p.89): for any discrete RV, we have:
  - $0 \leq p(y) \leq 1$ for all y
  - $\sum_y p(y) = 1$, where the sum here is over all values of y with nonzero probability
- p.90, ex.3.7: Y = # of empty bowls

  S = { aaa, aab, aac, aba, abb, abc, aca, acb, acc, baa, bab, bac, bba, bbb, bbc, bca, bcb, bcc, caa, cab, cac, cba, cbb, cbc, cca, ccb, ccc }

  From the listing of S, we see p(0) = 6/27 = 2/9, p(1) = 18/27 = 6/9 = 2/3, and p(2) = 3/27 = 1/9; table is:

| y | P(y) | y × P(y) | $y^2$ × P(y) |
|---|---|---|---|
| 0 | 2/9 | 0 | 0 |
| 1 | 6/9 | 6/9 | 6/9 |
| 2 | 1/9 | 2/9 | 4/9 |
| Total | 9/9 = 1 | 8/9 | 10/9 |

## Section 3.3

- **Def. 3.4** (p.91): Y is a discrete RV with probability function p(y), then the expected value of Y, E(Y) is defined to be $E(Y) = \sum_y yP(y)$(taken over all y with non-zero probability) – provided the sum is absolutely convergent, i.e., that $\sum_y |y|P(y) < \infty$

- **Theorem 3.2 (p.93): Y is a discrete RV with probability function p(y) and g(Y) is a real-valued function of Y, then the expected value of g(Y) is:**

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

- **Def. 3.5 (p.93). Y is a RV with mean E(Y) = μ, then the variance of Y is**

$$\sigma^2 = E\left[(Y-\mu)^2\right]$$

  **The standard deviation (σ) is the positive square root of $\sigma^2$. The units of μ and σ are the same as for y.**

- **Example 3.2. The table and histogram are given on p.94. We get μ = (0)(1/8) + (1)(2/8) + (2)(3/8) + (3)(2/8) = 14/8 = 7/4 = 1.75. Also, $\sigma^2 = (0 - 1.75)^2 (1/8) + (1 - 1.75)^2 (2/8) + (2 - 1.75)^2 (3/8) + (3 - 1.75)^2 (2/8) = 7.5/8 = 0.9375$, and so $\sigma = \sqrt{0.9375} = 0.9682$. Then, μ ± σ is (0.7818,2.7182), which contains 5/8 = 0.625 of the probability mass, close to the empirical rule (68%)**

- **Theorems 3.3 – 3.5 (p.95) can be combined to say that for Y a discrete RV with probability function p(y) and a,b,c are constants, then E(•) is a linear operator:**

$$E[af(Y) + bg(Y) + c] = aE[f(Y)] + bE[g(Y)] + c$$

- **Theorem 3.6 (p.96). Let Y be a RV with probability function p(y) and mean E(Y) = μ, then**

$$\sigma^2 = E\left[Y^2\right] - \mu^2$$

  **(sometimes called the *short-cut formula* for $\sigma^2$)**

- Example 3.2 continued. $E(Y^2) = (0^2)(1/8) + (1^2)(2/8) + (2^2)(3/8) + (3^2)(2/8) = 32/8 = 4$, so $\sigma^2 = 4 - 1.75^2 = 15/16$
- p.90, ex.3.7 continued. $\mu = E(Y) = 8/9$, $E(Y^2) = 10/9$ so $\sigma^2 = 10/9 - (8/9)^2 = 26/81 = 0.3210 = 0.5666^2$
- p.99, ex.3.23 – the table and calculations are as follows

| y | P(y) | y × P(y) | $y^2 \times P(y)$ |
|---|---|---|---|
| -4 | 9/13 | -36/13 | 144/13 |
| 5 | 2/13 | 10/13 | 50/13 |
| 15 | 2/13 | 30/13 | 450/13 |
| | | $\mu = 4/13$ | $E(Y^2) = 644/13$ |

Here, $\sigma^2 = 644/13 - (4/13)^2 = 8356/169 = 7.0316^2$. So, each time you play, you expect to win \$4/13 = 0.3077 = about 31¢ give or take about \$7.

## Section 3.4

- Def. 3.6 (p.101). A binomial experiment has *all* of the following properties:
  1. The experiment consists of a fixed number n of trials
  2. Each trial results in either a "success" S or "failure" F
  3. The success probability p stays the same for all trials
  4. The trials are independent
  5. Y, the random variable, is the number of successes observed in the n trials

- **Notice that just one violation above makes the RV not binomial: 3 draws *without replacement* from a hat containing 5 red chips and 10 green ones and counting the number of reds is not binomial. If sampling is done *with replacement*, then the RV would be binomial with n = 3 and p = ⅓; if we called a green chip a "success", this would be binomial with n = 3 and p = ⅔. Also, note that sampling until you get the $5^{th}$ red chip is not binomial since n is not set *a priori*.**
- **Def. 3.7 (p.103). The binomial RV Y based on n trials and with success probability p has the binomial probability distribution formula:**

$$p(y) = \binom{n}{y} p^y (1-p)^{n-y}, y = 0, 1 \dots n; 0 \le p \le 1$$

- **Plots of the prob. histograms are given on p.104 of Bin(n=10,p=0.1), Bin(n=10,p=½) and Bin(n=20,p=½)**
- **Letting q = 1 – p, notice that we get these probabilities from the binomial expansion:**

$$1 = 1^n = (q+p)^n = \binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} + \binom{n}{2} p^2 q^{n-2} + \cdots + \binom{n}{n-1} p^{n-1} q^1 + \binom{n}{n} p^n q^0 = P(0) + P(1) + \cdots + P(n-1) + P(n)$$

- **p. 105, ex. 3.7. In a lot of 5000 electrical fuses, 5% are bad, we take a sample of 5 fuses, and the probability of observing at least one defective is *approximately***

$$1 - p(0) = 1 - \binom{5}{0}(0.05)^0(0.95)^5 = 0.2262$$

- **The above solution is only approximate since one binomial condition is not truly met (that p stays the same across the "draws" since sampling is without replacement), but the approximation is very close**
- **The above calculations in R:**

  **1-pbinom(0,5,0.05)**

  **0.2262191**
- **Had we taken a sample of n = 20 fuses above and wanted the probability of at least 4 defective:**

  **1-pbinom(3,20,0.05)**

  **0.01590153**
- **p. 105, ex. 3.8. p = 0.30 (recovery rate) and n = 10 (sample taken) if we see y = 9 recoveries, we could calculate: P(9) + P(10) = 0.000144 – that the probability of *what we saw or more extreme* is very small, this makes us doubt that p = 30%, and conclude that the new medication may have been significantly improved**
- **Theorem 3.7 (p.107). Let Y be a binomial RV with parameters n (trials) and p (success probability). Then $\mu = E(Y) = np$ and $\sigma^2 = V(Y) = npq$ for q = 1 - p.**
- **Proofs of the above results are instructive and should be carefully worked through.**
- **p. 105, ex. 3.8. In a binomial setting with n = 20, we**

observe y = 6 “successes” (whether the employee favors the new retirement policy). To estimate p, we can use the technique of maximum likelihood (ML) estimation: the likelihood here is

$$L(p) = p(y) = \binom{n}{y} p^y (1-p)^{n-y} \propto p^6 (1-p)^{14}$$

Maximizing the likelihood is equivalent to maximizing the log-likelihood – whence,

$$\frac{d}{dp}[6ln(p) + 14ln(1-p)] = \frac{6}{p} - \frac{14}{1-p}$$

So the ML estimate (MLE) is $\hat{p} = \frac{6}{20} = 0.30$

## Section 3.5

- Related to the binomial distribution is the geometric distribution, where Y = # of ‘tosses’ for the first success
- Def. 3.8 (p.115). A RV Y has the Geometric probability distribution with success probability p if and only if

$$p(y) = q^{y-1}p \text{ for } y = 1, 2, 3 \dots, 0 \le p \le 1$$

- The probability histogram for a geometric RV with p = ½ on p.115
- p. 116, ex. 3.11. A “success” is engine malfunction during a one-hour period, p = 0.02, & we want the probability the engine survives two hours = P(Y ≥ 3) = 1 – P(Y ≤ 2) = 1 – p – qp = 1 – 0.02 – (0.02)(0.98) = 0.9604 Y = the # of one-hour intervals until the 1st malfunction

- Some other texts (and R) define the geometric RV to be the ***number of failures until the first success*** (Y*); noting that Y* = Y – 1, in the above, we have P(Y ≥ 3) = P(Y* ≥ 2) = 1 – P(Y* ≤ 1); in R, we get:

```
1-pgeom(1,0.02)
0.9604
```

- **Theorem 3.8** (p.116). Y has the geometric distribution (i.e., our definition!) with success p, then

$$\mu = E(Y) = 1/p \text{ and } \sigma^2 = V(Y) = \frac{1-p}{p^2}$$

- Again, the proofs are very instructive (need to use your geometric series, which are very important!)
- p. 116, ex. 3.11 (continued). For the engine failure ex., $\mu = 1/0.02 = 50$, $\sigma^2 = 0.98/0.02^2 = 2450 = 49.5^2$, so we expect to wait 50 hours give or take 50 hours.
- p.118, ex. 3.13. In a geometric setting with unknown p, the first person who likes the policy (success) is the 5th one interviewed, and we again use ML estimation:

$$L(p) = p(y) = (1-p)^{y-1}p = (1-p)^4 p$$

$$\frac{d}{dp}[4ln(1-p) + ln(p)] = \frac{-4}{1-p} + \frac{1}{p}$$

So, the ML estimate (MLE) is $\hat{p} = \frac{1}{5} = 0.20$

- Per p.119, ex. 3.71, for Y a Geometric RV with success probability p, we have (a) $P(Y > a) = q^a$, and so (b) the memory-less property: $P(Y > a + b | Y > a) = P(Y > b)$

## Section 3.6

- **Related to the GEO distribution is the Negative Binomial NB distribution which waits for the $r^{th}$ success**
- **For y ≥ r, if the $r^{th}$ success occurs on trial y, then we know that (r-1) successes occurred on trials 1 to (y-1), and this latter event is binomial – this leads to the following probability function for the NB distribution:**

$$p(y) = \binom{y-1}{r-1} p^r q^{y-r};\ y = r, r+1, r+2 \ldots, 0 \le p \le 1$$

- **<u>p.122, ex.3.14</u>: drilling oil wells with p = 0.2, r = 3, then** $p(5) = \binom{5-1}{3-1}(0.2)^3(0.8)^2$ **= 0.0307**
- **The R command for the previous calculation is "dnbinom($y_0$-r,r,p)" or here "dnbinom(2,3,0.2)"**
- **Not requested, but note that the R command "pnbinom(2,3,0.2)" is equivalent to "dnbinom(0,3,0.2) + dnbinom(1,3,0.2) + dnbinom(2,3,0.2)"**
- **For a NB random variable with parameters p and r,**

$$\mu = E(Y) = {}^{r}/_{p} \text{ and } \sigma^2 = V(Y) = \frac{r(1-p)}{p^2}$$

- **For <u>ex.3.14</u>, $\mu$ = 3/0.2 = 15 and $\sigma^2$ = 3(0.8)/$(0.2)^2$ = 60.**

## Section 3.7

- **The Hypergeometric HG distribution can be thought of as a 'sampling without replacement' analog of the Binomial distribution; the population size is N, and r are of one type (A) and the remaining (N-r) the other** type (B). Thus, the proportion of the type A objects is p = r/N, and we take a sample (without replacement) of size n. Y is the number of type A objects.

- The probability distribution for a HG RV (Y) is

$$p(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

- Constraints on y above are: $y \geq n+r-N$ and $y \leq r$
- <u>p. 126, ex.3.16</u>: N = 20, n = 10, r = y = 5,

$$p(5) = \frac{\binom{5}{5}\binom{15}{5}}{\binom{20}{10}} = 0.0163$$

- In R, instead of "choose(15,5)/choose(20,10)", we can just use "dhyper($y_0$,r,N-r,n)" - here "dhyper(5,5,15,10)"
- Y is a HG RV, then the mean and variance are

$$\mu = E(Y) = \frac{nr}{N} \text{ and } \sigma^2 = V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

- Note that $\mu = np$, and denoting the factor $\frac{N-n}{N-1}$ as $\phi$, we get $\sigma^2 = np(1-p)\phi$, similar to a BIN RV
- For given n, as $N \to \infty$, we get $\phi \to 1$, whence both the mean and variance coincide with BIN dist.; so too do the probability functions coincide since $N \to \infty$ means an infinite population in which case sampling with and without replacement are essentially the same.

- p.127, ex.3.17: N = 20, n = 5, p = 4/20 = 0.2, and we reject if Y > 1, so p(Y > 1) = 1 – p(0) – p(1) = 0.2487; also, $\mu$ = 5*0.20 = 1, $\sigma^2$ = 5*0.20*0.80*(15/19) = 0.6316
- In R: “1-phyper(1,4,16,5)” yields “0.24871”

## Section 3.8

- The Poisson POI distribution is related to the BIN distribution as derived on p.131: for the number of occurrences of some event over a time interval, we break the interval to n equal-length sub-intervals so that occurrences on the sub-intervals are independent and p(0) = 1-p, p(1) = p, and p(2 or more) = 0
- Here, p is the probability of an occurrence (a ‘success’) on any sub-interval
- Then the total number of occurrences on the larger interval has the POI distribution with mean $\lambda$ = np
- The POI probability function with parameter $\lambda$ is

$$p(y) = \frac{\lambda^y}{y!} e^{-\lambda} \text{ for } y = 0, 1, 2 \dots \text{ and } \lambda > 0$$

- In our derivations, recall that $\lim_{n\to\infty}\left(1 + \frac{k}{n}\right)^n = e^k$ and that $e^k = \sum_{z=0}^{\infty} \frac{k^z}{z!} = 1 + \frac{k}{1} + \frac{k^2}{2} + \frac{k^3}{6} + \frac{k^4}{24} + \cdots$
- For a given problem or exercise, always note the [larger] interval length (see ex. 3.22 below)

- **p.132, ex.3.19**: Per 30-minute period, $\lambda$ = 1 (one visit per half-hour period), so $p(y) = \frac{1}{y!}e^{-1}$ and p(0) = $e^{-1}$/1 = 0.368, p(1) = $e^{-1}$/1 = 0.368, and p(2) = $e^{-1}$/2 = 0.1839; P(Y ≥ 1) = 1 - p(0) = 1 – $e^{-1}$ = 0.6321
- In R, "dpois(2,1)" gives "0.1839397", and for P(Y ≥ 1), typing "1-ppois(0,1)" yields "0.6321206"
- **p.134, ex.3.21** demonstrates that the POI distribution can provide a good approximation to the BIN ***for large n & small p***: the exact BIN answer is pbinom(3,20,0.1) = 0.8670467, and POI gives ppois(3,2) = 0.8571235.
- p. 134, **Theorem 3.11** states that for Y a POI RV with parameter $\lambda$, $\mu = \sigma^2 = \lambda$. Again, proofs are important.
- **p.135, ex.3.22** (Poisson Process): average is stated as 3 accidents per month, but since the rest of the exercise concerns 2-month period, take $\lambda$ = 2(3) = 6 (accidents per 2-month period). We obtain the desired P(Y≥10) using R: 1-ppois(9,6) = 0.08392402. Since this p-value is not unusually small, we can conclude that 10 accidents in the past 2 months is not indicative that the mean has increased (from old mean of $\lambda$ = 6).

## Section 3.9

- We define a function, called the moment generating function, MGF, which can be used to obtain (generate) the moments of a distribution. At this point, the MGF appears only theoretical, but we will find it very useful later to identify the distribution of a given RV.

- p.138, Definition 3.12: The k^th moment of the RV Y ***about the origin*** or ***about zero*** is $\mu_k' = E\left(Y^k\right)$
- p.138, Definition 3.13: The k^th moment of Y ***about its mean*** or ***central moment*** is $\mu_k = E\left[(Y-\mu)^k\right]$
- Thus, $\mu_1' = \mu$, $\mu_2' = E\left(Y^2\right) = \sigma^2 + \mu^2$, $\mu_1 = 0$, and $\mu_2 = \sigma^2$; other moments (such as skewness and kurtosis) can also be calculated.
- p.139, Definition 3.14: The MGF, m(t), for a RV Y is $m(t) = E(e^{tY})$ provided there exists a positive constant b such that m(t) is finite for |t| ≤ b
- Provided m(t) exists, it's easy to show that (see p.139)

$$m(t) = E\left(e^{tY}\right) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots$$

- p.139, Theorem 3.12: If m(t) exists, then for any positive integer k,

$$\left.\frac{d^k m(t)}{dt^k}\right|_{t=0} = m^{(k)}(0) = \mu_k'$$

That is, the k^th derivative of m(t) with respect to t and evaluated at t = 0 gives $\mu_k'$.

- It is shown (p.140, ex.3.23; p.142, ex.3.145; and p.142, ex.3.147) that the MGFs for the Poisson, Binomial and Geometric distributions are as follows (center column):

| Distribution | MGF | PGF |
|---|---|---|
| Poisson | $m(t) = e^{\lambda(e^t-1)}$ | p.146 ex.3.165 |
| Binomial | $m(t) = (pe^t + q)^n$ | p.146 ex.3.164 |
| Geometric | $m(t) = \dfrac{pe^t}{1-qe^t}$ | $\pi(t) = \dfrac{pt}{1-qt}$ |

- **For the GEO MGF, it is necessary that $qe^t < 1$, i.e., that $t < -\ln(q)$. Since there is an interval around zero for which $E(e^{tY})$ exists, m(t) is indeed well-defined.**
- **<u>p.141, ex.3.24</u>: For a Poisson RV, $m(t) = e^{\lambda(e^t-1)}$, so $m'(t) = \lambda e^t e^{\lambda(e^t-1)}$ so $\mu = m'(0) = \lambda$ and $m''(t) = \lambda\left[\lambda e^{2t} e^{\lambda(e^t-1)} + e^t e^{\lambda(e^t-1)}\right]$ thus $\mu_2' = m''(0) = \lambda(\lambda+1) = \lambda^2 + \lambda$. So, $\sigma^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.**
- **<u>p.141, ex.3.25</u>: If Y is a RV with MGF $m(t) = e^{3.2(e^t-1)}$, then we know it has a Poisson distribution with mean $\lambda = 3.2$. That is, we can identify the distribution by the *<u>uniqueness property</u>* of MGFs.**

## <u>Section 3.10</u>

- **p.144, <u>Definition 3.15</u>: For the discrete RV Y, let P(Y=k) be denoted $p_k$ for k=0,1,2,…, then for all t such that it is finite, the probability generating function PGF of Y is**

$$\pi(t) = E(t^Y) = \sum_{k=0}^{\infty} p_k t^k$$

- p.144, Definition 3.16: For the RV Y and k a positive integer, the $k^{th}$ factorial moment is

$$\mu_{[k]} = E\left(Y^{(k)}\right) = E[Y(Y-1)(Y-2)\ldots(Y-k+1)]$$

- p.144, Theorem 3.13: If $\pi(t)$ is the PGF for the integer-valued RV Y, then we can obtain factorial moments by:

$$\left.\frac{d^k \pi(t)}{dt^k}\right|_{t=1} = \pi^{(k)}(1) = \mu_{[k]}$$

- p.145, ex.3.26: For GEO and t < 1/q, $\pi(t) = \frac{pt}{1-qt}$
- p.145, ex.3.27: From above, $\pi'(t) = p(1-qt)^{-2}$, so $\pi'(1) = p/(1-q)^2 = 1/p$. Also, $\pi''(t) = 2pq(1-qt)^{-3}$, so $\mu_{[2]} = 2q/p^2$

## Section 3.11

- Empirical Rule: For distributions that resemble the Normal distribution, approximately 68%, 95% and 99.7% are within 1, 2 and 3 $\sigma$'s of $\mu$.
- For ***any distribution***, however, Chebyshev's theorem gives a lower bound: for constant k > 0,

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

Equivalent to the above is:

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

- Thus, for any distribution, ***lower bounds*** to the coverage within 1, 2 and 3 $\sigma$'s of $\mu$ are: 0, 75% & 88.9%