

Probability Theory

Chapter 4—Random variables

Nan Liu¹

¹These slides are based on Prof. Mui Li's. Any mistakes are my own. If you find an error, please send an email at nanliu@xmu.edu.cn

4.1 Random Variables

- In many scenarios, we are interested in a function of the outcome as opposed to the actual outcome; e.g. we are interested in the sum of two dice and not in the separate values of each die or, when we flip a coin, we want to know the number of tails.
- Real-valued functions defined on the sample space are *random variables*; e.g. number of gold medals by Canadian athletes at Olympics, your score at the SAT test etc.

4.1 Examples

- **Toss Coin:** For a coin toss, the possible events are heads or tails. The number of heads appearing in one fair coin toss can be described using the following random variable:

$$X = \begin{cases} 0 & \text{if head} \\ 1 & \text{if tail} \end{cases} \quad \text{and } P(X = 0) = P(X = 1) = \frac{1}{2}$$

Remark: Note that we could define completely arbitrary random variable such as $X = 10$ if heads and $X = \pi$ if tail. This is of no “practical” interest but possible conceptually.

- **Toss 3 coins:** We are tossing 3 times a fair coin. If we call X the number of heads obtained, then it is a random variable such that

$$\begin{aligned} P\{Y = 0\} &= P(\{T, T, T\}) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}, \\ P\{Y = 1\} &= P(\{T, T, H\}, \{T, H, T\}, \{H, T, T\}) = \frac{3}{8}, \\ P\{Y = 2\} &= P(\{T, H, H\}, \{H, T, H\}, \{H, H, T\}) = \frac{3}{8}, \\ P\{Y = 3\} &= P(\{H, H, H\}) = \frac{1}{8}. \end{aligned}$$

4.1 Examples

- **Rolling a fair die:** A random variable can also be used to describe the process of rolling a fair dice and the possible outcomes. The most obvious representation is to take the set $\{1, 2, 3, 4, 5, 6\}$ as the sample space, defining the random variable X as the number rolled. In this case,

$$X = i \text{ if a } i \text{ is rolled and } P(X = i) = \frac{1}{6} \text{ for } i = 1, \dots, 6.$$

4.1 Example

- Assume you have trials consisting of sitting an exam. You have probability p of passing the exam. If you fail, you sit the exam again. These trials are considered independent as you never both revise the material. We denote X the number of exams you need to sit before passing the subject.
- We have

$$P(X = 1) = P(\text{Pass}) = p,$$

$$P(X = 2) = P(\text{Fail}, \text{Pass}) = (1 - p)p,$$

....

$$P(X = n) = P(\underbrace{\text{Fail}, \dots, \text{Fail}}_{n-1 \text{ times}}, \text{Pass}) = (1 - p)^{n-1} p.$$

- You can check that

$$\sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} (1 - p)^{k-1} p = 1.$$

4.2 Discrete Random Variables

- A random variable (r.v.) X which can take at most a countable number of possible values is said to be discrete.
- The *probability mass function (pmf)* of X is denoted by

$$p(a) = P(X = a).$$

- If X can take the values $\{x_1, x_2, \dots, \}$, then

$$p(x_i) \geq 0 \quad \text{and} \quad \sum_{i=1}^{\infty} p(x_i) = 1.$$

- *Cumulative distribution function (CDF)* is:

$$F(a) = P(X \leq a) = \sum_{x: x \leq a} p(x).$$

4.2 Examples of discrete random variables

- If a r.v. X has a pmf given by

$$p(1) = 1/4, p(2) = 1/2, p(3) = 1/8, p(4) = 1/8$$

find its CDF and plot it.

- The probability mass function (pmf) of a r.v. X is given by

$$p(i) = c\lambda^i/i!, i = 0, 1, 2, \dots,$$

where λ is some positive value. Find (a) $P(X = 0)$ and (b) $P(X > 2)$.

4.3 Expected Value

- The expectation of X , denoted $E(X)$, is defined as

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i) = \sum_{x:p(x)>0} xp(x)$$

- It corresponds to a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it.
- **Example:** Assume you have a fair die and X is the number rolled then $p(i) = 1/6$ for $i = 1, 2, \dots, 6$ and

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \dots + 6 \times \frac{1}{6} = \frac{7}{2}$$

4.3 Examples

- You fancy a person but you feel that he/she might be out of your league. Your utility is your happiness whose reference level is zero. You'd like to maximize your average happiness. If you could date this person, then your happiness would increase by say 1, 000. If you ask her/him out and you got rejected, then your happiness will plunge to -100. Given the probability that the person accepts going out with you is 0.1, should you try to ask her/him out?
- A school class of 120 students is driven in 3 buses to a symphonic performance. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, and find $E[X]$.

4.4 Expectations of a function of a random variable

- Given a random variable X , we can define a new random variable $Y = g(X)$. From the probability mass function (pmf) of X , we can obtain the pmf of Y .
- **Example:** Assume you have X taking values in $\{-2, -1, 0, 1, 2\}$ with equal probability and consider $Y = X^2$. Calculate EY .

4.4 Expectations of a function of a random variable

- Let X be a discrete r.v. taking values $\{x_i\}$ with pmf $p_X(x_i)$ and define $Y = g(X)$.
- **Proposition.** We have

$$E(Y) = E(g(X)) = \sum_i g(x_i) p_X(x_i).$$

- **Corollary** If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

4.4 Example: Optimal Stock

- Bingdundun is currently sold at a net profit of $b\$$ for each unit sold and will bring a net loss of $l\$$ for each unit left unsold after the Olympics. The number of units of the product sold at a specific store is a random variable X with pmf $p(i)$, $i \geq 0$. If the store must stock this product in advance, compute the expected profit for a determined number of units s . Find $\hat{s}$ which optimizes the maximum expected profit.
- The profit is equal to

$$\begin{aligned} Y(s) &= bX - l(s - X) \text{ if } X \leq s \\ &= sb \quad \text{if } X > s \end{aligned}$$

- Hence the expected profit is

$$\begin{aligned} E[Y(s)] &= \sum_{i=0}^s (bi - l(s - i)) p(i) + \sum_{i=s+1}^{\infty} sbp(i) \\ &= sb + (b + l) \sum_{i=0}^s (i - s) p(i) \end{aligned}$$

4.4 Example: Optimal Stock

- We want to maximize the profit. We note that

$$E[Y(s+1)] - E[Y(s)] = b - (b+l) \sum_{i=0}^s p(i)$$

so we have

$$E[Y(s+1)] - E[Y(s)] \geq 0$$

is equivalent to

$$\sum_{i=0}^s p(i) < \frac{b}{b+l}.$$

- So we should consider s^* the largest value such that $\sum_{i=0}^{s^*} p(i) < \frac{b}{b+l}$ then

$$E(Y(0)) < \dots < E(Y(s^*)) < E(Y(s^*+1)) > E(Y(s^*+2)) > \dots$$

so the optimal stock is $\hat{s} = s^* + 1$.

4.5 Variance

- A r.v. X is entirely defined by its p.m.f. but it is very useful to be able to summarize the essential properties of it through a few suitably defined measures.
- We have defined previously the expected value $E(X)$. However this measure tells us nothing about the variations of X ; e.g. consider $X = 0$ with proba. 1, $Y = 1$ w.p. 0.5 and $Y = -1$ w.p. 0.5 and finally $Z = 100$ w.p. 0.5 and $Z = -100$ w.p. 0.5 then

$$E(X) = E(Y) = E(Z) = 0$$

whereas there is a much greater spread/variability in the values of Y than in the deterministic value of X and the variability of Z than Y .

- We expect X to be around its expected value/mean $\mu = E(X)$ so it would make sense to quantify the average deviations from it. This is what the variance achieves.

4.5 Variance

- If X is a r.v. with expected value/mean $\mu = E(X)$ then the variance is

$$\text{Var}(X) = E\left((X - \mu)^2\right)$$

and $\text{Std}(X) = \sqrt{\text{Var}(X)}$ is called standard deviation.

- We have

$$\text{Var}(X) = E(X^2) - \mu^2$$

- **Example:** Consider a fair die with $X = i$ is “number i rolled” then we have seen that $E(X) = \frac{7}{2}$ and

$$\begin{aligned} E(X^2) &= 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} \\ &= \frac{1 + 4 + 9 + 16 + 25 + 36}{6} = \frac{91}{6} \end{aligned}$$

so

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

4.5 Variance

- Consider $Y = aX + b$ then

$$\text{Var}(Y) = a^2 \text{Var}(X)$$

- We have

$$\begin{aligned}\text{Var}(Y) &= E\left((Y - E(Y))^2\right) \\&= E\left[\left(aX + b - \underbrace{(aE(X) + b)}_{=E(Y)}\right)^2\right] \\&= E\left[a(X - E(X))^2\right] \\&= a^2 E\left[(X - E(X))^2\right] = a^2 \text{Var}(X)\end{aligned}$$

4.6 Bernoulli Distribution

- Assume you have an experiment which can be classified as either success or failure.
- If you associate a random variable X such that $X = 1$ corresponds to a success and $X = 0$ corresponds to a failure then

$$P(X = 1) = p(1) = p$$

$$P(X = 0) = p(0) = 1 - p.$$

- Such a random variable is called a *Bernoulli* random variable.
- We have

$$E(X) = p \text{ and } \text{Var}(X) = p(1 - p).$$

4.6 Binomial Distribution

- Consider a sequence of n independent success/failure experiments, each of which yields success with probability p .
- Let X denote the number of successes then we have $X = X_1 + X_2 + \cdots + X_n$ where X_i is a Bernoulli random variable with proba. p . It is a binomial random variable of parameters (n, p) .
- X is a discrete random variables in $\{0, 1, \dots, n\}$ such that

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- This expression follows from the fact that any specific sequence with exactly k successes (hence $n - k$ failures) has a proba $p^k (1 - p)^{n-k}$. There are $\binom{n}{k}$ such sequences.

4.6 Example

- *Example:* Everyday one of your classmates is taking the tram twice without any ticket. There is a 5% chance of meeting inspectors each time one boards the tram. What is the proba that he is going to be caught at least twice in a 30 day month?
- *Answer:* Let X be the number of times he will meet inspectors per year. X is a binomial random variable of parameters $n = 2 \times 30$, $p = 0.05$ and

$$\begin{aligned}P(X \leq 1) &= P(X = 0) + P(X = 1) \\&= \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1} \\&= (1-p)^n + np(1-p)^{n-1} \\&= 0.0461 + 0.1455 = 0.1916\end{aligned}$$

so

$$P(X \geq 2) = 1 - P(X \leq 1) = 0.8084$$

4.6 Example

- *Example:* Your friend pretends that he is able to distinguish pepsi from coke. You give him a test. Out of 5 trials, he gets the right answer 4 times. What is the probability that he just got lucky?
- *Answer:* Let X be the number of right answers. If he answers completely randomly, then X is a binomial of parameters $n = 5$ and $p = \frac{1}{2}$. We have

$$\begin{aligned}P(X \geq 4) &= P(X = 4) + P(X = 5) \\&= \binom{5}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{5-4} + \binom{5}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{5-0} \\&= 0.1875.\end{aligned}$$

This is no convincing evidence that your friend can distinguish the two beverages.

4.6 Example

- *Example:* A survey from Teenage Research Unlimited (Northbrook, Ill.) found that 30% of teenage consumers receive their spending money from part-time jobs. If five teenagers are selected at random, find the probability that at least three of them will have part-time jobs.
- *Answer:* Let X be the number of teenagers having a part-time job among 5 teenagers, then X is Binomial of parameters $n = 5$, $p = 0.3$ so

$$\begin{aligned}P(X \geq 3) &= P(X = 3) + P(X = 4) + P(X = 5) \\&= 0.132 + 0.028 + 0.002 = 0.162\end{aligned}$$

- *Example:* What is the probability of obtaining 45 or fewer heads in 100 tosses of a fair coin?
- *Solution:* Let X be the number of heads, then X is binomial of parameters $n = 100$, $p = 0.5$ and

$$P(X \leq 45) = \sum_{k=0}^{45} \binom{n}{k} p^k (1-p)^{n-k} = 0.184.$$

4.6 Mean of the Binomial Random Variable

- The mean/expected value of X is given by

$$E(X) = np$$

- Proof.** We have

$$\begin{aligned} E(X) &= \sum_{k=0}^n k P(X = k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k \frac{n(n-1)!}{k(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{l=0}^{n-1} \frac{(n-1)!}{l!(n-1-l)!} p^l (1-p)^{n-1-l} \quad (m \leftarrow n-1, l \leftarrow k-1) \\ &= np \end{aligned}$$

- A much more straightforward proof follows from $X = X_1 + X_2 + \cdots + X_n$ where X_i is a Bernoulli random variable with proba. p so

$$\begin{aligned} E(X) &= E(X_1 + \cdots + X_n) \\ &= E(X_1) + \cdots + E(X_n) = np. \end{aligned}$$

4.6 Variance of the Binomial Random Variable

- The variance of the binomial random variable is

$$\text{Var}(X) = np(1-p).$$

- Proof 1.** We have $\text{Var}(X) = E(X^2) - E(X)^2 = E(X^2) - (np)^2$ where

$$\begin{aligned} X^2 &= (X_1 + \dots + X_n)^2 \\ &= \sum_{k=1}^n X_k^2 + \sum_{k,l=1; k \neq l}^n X_k X_l \end{aligned}$$

so

$$E(X^2) = \sum_{k=1}^n E(X_k^2) + \sum_{k,l=1; k \neq l}^n E(X_k X_l).$$

We have $E(X_k^2) = 1^2 \times p + 0^2 \times (1-p) = p$. To compute $E(X_k X_l)$ where $k \neq l$, we note that $X_k X_l \in \{0, 1\}$ is a r.v. such that

$$\begin{aligned} P(X_k X_l = 1) &= P(X_k = 1) P(X_l = 1) = p^2, \\ P(X_k X_l = 0) &= 1 - P(X_k X_l = 1) = 1 - p^2. \end{aligned}$$

4.6 Variance of the Binomial Random Variable

- As a safety check, you have indeed

$$\begin{aligned}P(X_k X_l = 0) &= P(X_k = 0, X_l = 0) + P(X_k = 0, X_l = 1) \\&\quad + P(X_k = 1, X_l = 0) \\&= (1-p)^2 + (1-p)p + p(1-p) \\&= 1 - p^2.\end{aligned}$$

- Hence we have

$$E(X_k X_l) = 1 \times p^2 + 0 \times (1 - p^2) = p^2$$

thus

$$\begin{aligned}E(X^2) &= \sum_{k=1}^n E(X_k^2) + \sum_{k=1, l > k}^n E(X_k X_l) \\&= n \times p + n(n-1) \times p^2\end{aligned}$$

- Finally we have

$$\text{Var}(X) = np + n(n-1)p^2 - (np)^2$$

4.6 Variance of the Binomial Random Variable

- **Proof 2.** We have $\text{Var}(X) = E(X^2) - E(X)^2$ where

$$\begin{aligned} E(X^2) &= \sum_{k=0}^n k^2 P(X=k) = \sum_{k=1}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n k n \binom{n-1}{k-1} p p^{k-1} (1-p)^{n-k} \\ &= \sum_{k=1}^n k n \binom{n-1}{k-1} p p^{k-1} (1-p)^{n-k} = np \sum_{k=1}^n k \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{l=0}^{n-1} (l+1) \frac{(n-1)!}{l! (n-1-l)!} p^l (1-p)^{n-1-l} = np E[Y+1] \end{aligned}$$

where Y is Binomial $(n-1, p)$ so $E[Y+1] = (n-1)p + 1$. Finally $\text{Var}(X) = np[(n-1)p + 1] - (np)^2 = np(1-p)$.

4.6 Example: Blood test

- A large number, N , of people are subjected to a blood test. This can be administered in two ways: (1) Each person can be tested separately, in this case N tests are required, (2) the blood samples of k persons can be pooled and analyzed together. If this test is negative, this one test suffices for the k people. If the test is positive, each of the k persons must be tested separately, and in all, $k + 1$ tests are required for the k people. Assume that the probability p that a test is positive is the same for all people and that these events are independent.
- (a) Find the probability that the test for a pooled sample of k people will be positive.
- (b) What is the expected value and variance of the number Y of tests necessary under plan (2)? (Assume that N is divisible by k .)

4.6 Example: Blood test

- (a) We have

$$\begin{aligned}P(\text{pooled sample } +) &= P(\text{at least one of the } k \text{ people } +) \\&= 1 - P(\text{no people } +) = 1 - (1 - p)^k.\end{aligned}$$

- (b) We have $M = N/k$ pooled samples. Let X_i be a Bernoulli variable of parameter $q = 1 - (1 - p)^k$ corresponding to success="pooled sample +" then

$$X_i = \begin{cases} 1, & \text{w.p. } (1 - p)^k \\ k + 1, & \text{w.p. } 1 - (1 - p)^k \end{cases}$$

where $Y = X_1 + X_2 + \cdots + X_M$ is a Binomial of parameters (M, q) . So we have

$$E(Y) = \sum_{i=1}^M E(X_i) = M\{k + 1 - k(1 - p)^k\},$$

$$\text{Var}(Y) = \sum_{i=1}^M \text{Var}(X_i) = ?.$$

4.7 Poisson Random Variable

- A discrete r.v. X taking values $0, 1, 2, \dots$ is said to be a Poisson r.v. with parameter λ , $\lambda > 0$, if

$$p(i) = P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}.$$

- This is indeed a probability distribution as

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{i!}}_{=e^{\lambda}} = 1.$$

- This expresses the probability of a number of events occurring in a fixed period of time if these events occur with a *known average rate* λ .
- This is used absolutely everywhere as many natural phenomena are Poisson distributed: the number of phone calls at a call center per minute, the number of murders in London etc.

4.7 Example: Misprints in a textbook

- *Example:* There are 50 misprints in a book which has 250 pages and assume these errors follow a Poisson distribution of parameter $\lambda = 50/250 = 0.2$. Find the probability that page 100 has no misprints. Find the probability that page 100 has 2 misprints.
- *Answer:* If we let X be the r.v. denoting the number of misprints on page 100 (or any other page), X is a Poisson r.v. with parameter $\lambda = 0.2$. So we have

$$\text{Proba no misprint} = P(X = 0) = e^{-0.2} \frac{(0.2)^0}{0!} = e^{-0.2} = 0.819,$$

$$\text{Proba 2 misprints} = P(X = 2) = e^{-0.2} \frac{(0.2)^2}{2!} = 0.0164.$$

4.7 Example: Waiting at the bus stop

- *Example:* Assume that 6 buses per hour stop at your bus stop. If the buses were arriving exactly every 10 minutes, then your average waiting time would be 5 minutes; i.e. you wait between 0 and 10 minutes. What is the probability that you are going to wait at least 5 minutes without seeing any bus if the buses follow a Poisson distribution? What is the proba to wait at least 10 minutes? What is the proba of seeing two buses in 10 minutes?
- *Answer:* If we let X_1 be the number of buses arriving in 5 minutes, it is a Poisson r.v. with parameter 0.5 (average rate 6 per hour). So we have

$$P(X_1 = 0) = e^{-0.5} = 0.60.$$

If we let X_2 be the number of buses arriving in 10 minutes, it is a Poisson r.v. with parameter 1. So we have

$$P(X_2 = 0) = e^{-1} = 0.368, \quad P(X_2 = 2) = e^{-1} \frac{1^2}{2!} = 0.184.$$

4.7 Example: Murders in London

- Each year the London Metropolitan Police record around 170 murders, and this has been stable for the last 5 years. Each of these murders is an individual crime that cannot be predicted. It may appear strange, but this very randomness means that the overall pattern of murders is in some ways quite predictable.
- Assuming the number of murders are Poisson distributed, compute the probability of having no murder during a day, 3 or more murders in one day, a week without any murder.

4.7 Example: Murders in London

- We have an average rate per day of $\lambda = 170/365$.
- *Peaceful day* (no murder). The proba that of having no murder is $P(X = 0) = e^{-\lambda} = 0.63$.
- *Bloody day* (3 or more murders). The proba is $P(X \geq 3) = 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2} e^{-\lambda} = 0.0014 \approx \frac{0.5}{365}$
- *Peaceful week* (no murder over a week). We have a Poisson distribution $\lambda' = 170/365 \times 7$

$$P(X = 0) = e^{-\lambda'} \approx 0.0384 \approx \frac{2}{52}.$$

4.7 Poisson as an approximation to Binomial

- If we consider a binomial r.v. X of parameters (n, p) such that n is large and p is small enough so that np is moderate then the binomial distribution can be well-approximated by the Poisson distribution of parameter $\lambda = np$.
- We have indeed for $\lambda = np \Leftrightarrow p = \frac{\lambda}{n}$

$$\begin{aligned}P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\&= \frac{n!}{k! (n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\&= \frac{n(n-1) \cdots (n-k+1)}{n^k} \frac{\lambda^k}{k!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k} \\&= 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{\lambda^k}{k!} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k}\end{aligned}$$

4.7 Poisson as an approximation to Binomial

- Now recall from your calculus course that

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}, \quad \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1,$$
$$\lim_{n \rightarrow \infty} \left(1 - \frac{j}{n}\right) = 1$$

so

$$\lim_{n \rightarrow \infty} P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

- Practically this means that if a large number of trials n are made which have a success probability p such that np is moderate then we can approximate the total number of successes by a Poisson r.v. of parameter $\lambda = np$.

4.7 Example: Sampling a population

- A certain disease occurs in 1.2% of the population. 100 people are selected at random from the population. This is a binomial experiment, where the number of trials 100 is large, and the expected number of people with the disease is $\mu = 100 \times 0.012 = 1.2$ is small.
- Assume we want to compute the probability of no people having the disease in a sample of 100. Let X being the number of people having the disease, then X is binomial of parameters $n = 100$ and $p = 1.2/100$ and

$$P(X = 0) = \binom{100}{0} \left(\frac{1.2}{100}\right)^0 \left(1 - \frac{1.2}{100}\right)^{100} = 0.299016$$

- We can *approximate* X by a Poisson random variable of parameter $\mu = 1.2$. We obtain in this case

$$P(X = 0) \approx e^{-\mu} \frac{\mu^0}{0!} = .301194$$

4.7 Example: Sampling a population

- The probability that exactly 2 people in the sample of 100 have the disease is following the binomial distribution

$$P(X = 2) = \binom{100}{2} \left(\frac{1.2}{100}\right)^2 \left(1 - \frac{1.2}{100}\right)^{98} = 0.2183$$

whereas the Poisson approximation gives

$$P(X = 2) \approx e^{-\mu} \frac{\mu^2}{2!} = 0.2169.$$

4.7 Example: Random Seeds

- You have a 10 meter by 10 meter plot of land, which is divided into a grid of 100 squares. You scatter 500 seeds on this plot. We assume that each seed falls at random, so that it is equally likely to fall on any of the 100 squares. Consider the square in the upper left hand corner. What is the probability that exactly 4 seeds fall on it? What is the probability that 0 seeds fall on it?
- Think of this as dropping 500 seeds, one after the other, and recording whether the seed falls into the upper left hand square or not. “Success” means falling into the square, and that happens with probability $1/100 = .01$. So, $n = 500$, and $p = 0.01$. The expected number of successes is $np = 5$. By the Poisson approximation, we have

$$P(\text{exactly 4 seeds}) \approx e^{-5} \frac{5^4}{4!} = 0.175.$$

The probability of no seeds is

$$P(\text{no seed}) \approx e^{-5} = 0.0067.$$

4.7 Example: Lunch Hour at restaurant

- During the noon lunch hour, 47 customers will walk through the door of the post office. Assume that each person arrives at a random time, independent of the other customers. What is the probability that more than one person walks through the door during the first minute?
- To see why this is binomial, think of each of the 47 persons choosing at random a minute during which to arrive. The probability that they choose the first minute is then $1/60$. Thus, we have 47 repetitions of a trial, where the probability of success is $1/60$. The expected number of successes is $np = 47/60 \approx 0.783$.
- To find the probability, we say

$$\begin{aligned}P(\text{more than 1 arrival}) &= 1 - P(\text{no arrival}) \\&\approx 1 - e^{-0.783} - 0.783e^{-0.783} \\&\approx 0.185.\end{aligned}$$

4.7 Mean of a Poisson Random Variable

- We have $E(X) = \lambda$.
- We have

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} i P(X = i) = \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} \\ &= e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda \lambda^{i-1}}{(i-1)!} \\ &= e^{-\lambda} \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \quad (\text{change } j \leftarrow i-1) \\ &= e^{-\lambda} \lambda e^{\lambda} = \lambda. \end{aligned}$$

- Note that this is in agreement with our approximation of Binomial by Poisson. A Binomial has mean np and we approximate it by a Poisson of parameter $\lambda = np$ which is also the mean of the Poisson distribution.

4.7 Variance of a Poisson Random Variable

- We have $\text{Var}(X) = E(X^2) - E(X)^2 = \lambda$.
- Proof: We have

$$\begin{aligned} E(X^2) &= \sum_{i=0}^{\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} = \sum_{i=1}^{\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} \\ &= e^{-\lambda} \sum_{i=1}^{\infty} i \frac{\lambda \lambda^{i-1}}{(i-1)!} = \lambda \sum_{j=0}^{\infty} (j+1) e^{-\lambda} \frac{\lambda^j}{j!} \\ &= \lambda E(Y+1) \end{aligned}$$

where Y is a Poisson random variable of parameter λ , hence $E(Y+1) = \lambda + 1$.

- We can now conclude

$$\text{Var}(X) = \lambda(\lambda + 1) - \lambda^2 = \lambda.$$

- Note that for Poisson random variable, we have $E(X) = \text{Var}(X) = \lambda$. For binomial we have $E(X) = np$ and $\text{Var}(X) = np(1-p)$ and so $\text{Var}(X) \approx E(X)$ only if $p \ll 1$.

4.8 Geometric Random Variable

- Consider yet again independent trials, each with success proba p . These trials are performed until a success occurs.
- Let X the number of trials required then $X \in \{1, 2, \dots\}$ then it follows a geometric p.m.f.

$$P(X = n) = (1 - p)^{n-1} p.$$

- It is indeed a valid p.m.f. as

$$\sum_{k=1}^{\infty} (1 - p)^{k-1} p = p \sum_{l=0}^{\infty} (1 - p)^l = p \frac{1}{1 - (1 - p)} = 1.$$

4.8 Properties of geometric distribution

- The geometric distribution is the simplest of the **waiting time** distributions. The random variable X can be interpreted as the number of trials required to obtain the first success, so we are “waiting for a success” .
- The geometric distribution has the so-called “**memoryless**” property in the sense that for integers $s > t$, we have

$$P(X > s | X > t) = P(X > s - t)$$

4.8 Mean and Variance of a Geometric Random Variable

- We have

$$E(X) = \frac{1}{p} \text{ and } \text{Var}(X) = \frac{1-p}{p^2}.$$

- Proof of expression of $E(X)$: If we set $q = 1 - p$, we have

$$E(X) = \sum_{k=1}^{\infty} kq^{k-1}p = p \sum_{k=1}^{\infty} kq^{k-1}.$$

Now we know that

$$J(q) = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

so by taking the derivative with respect to q

$$J'(q) = \sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2} = \frac{1}{p^2}$$

so

$$E(X) = \frac{1}{p}.$$

4.8 Example: Collecting pictures

- **Question:** Assume that, every time your little brother buys a box of Wheaties, he receives a picture of one of the n players of the Canadian Hockey team. Let X_k be the number of additional boxes he has to buy, after he has obtained $k - 1$ different pictures, in order to obtain the next new picture. Thus $X_1 = 1$, X_2 is the number of boxes bought after this to obtain a picture different from the 1st pictured obtained, and so forth. a) What is the distribution of X_k ? (We assume that Wheaties does not favour any player, i.e. $\text{Proba}(\text{finding player } i) = \text{Proba}(\text{finding player } j) = 1/n$). Let Y the total number of boxes you must buy to get the n different players. b) What is the expectation of Y ?

4.8 Example: Collecting pictures

- **Answer a):** After having obtained $k - 1$ different pictures, a successful event corresponds to buying a box of Wheaties with a picture of one of the remaining $(n - (k - 1)) = (n - k + 1)$ player. The probability of this event is

$$p_k := (n - k + 1) \times \frac{1}{n} = 1 - \frac{(k - 1)}{n}.$$

Hence X_k follows a geometric distribution with p_k .

- **Answer b):** We have $Y = X_1 + X_2 + \cdots + X_n$ so

$$\begin{aligned} E(Y) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= \frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_n} \\ &= 1 + \frac{1}{1 - \frac{1}{n}} + \cdots + \frac{1}{1 - \frac{n-1}{n}} \end{aligned}$$

For $n = 6$, we have $E(Y) = 14.7$ and for $n = 14$, we have $E(Y) = 45.5$.

4.8 Negative Binomial Random Variable

- Suppose that independent trials, each having probability p , $0 < p < 1$, of being a success are performed until a total of r successes is accumulated.
- If we let X equal the number of trials required, then

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad n = r, r+1, \dots$$

- We can prove analytically that

$$\sum_{n=r}^{\infty} P\{X = n\} = \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = 1$$

4.8 Negative Binomial Random Variable

- The number of trials required to obtain r successes can be expressed

$$Y_1 + Y_2 + \cdots + Y_r,$$

where Y_1 equals the number of trials required for the first success, Y_2 the number of additional trials after the first success until the second success occurs, and so on.

- Note that a geometric random variable is just a negative binomial with parameter $(1, p)$.

4.8 Example

- The Banach match problem

At all times, a pipe-smoking mathematician carries 2 matchboxes—1 in his left-handpocket and 1 in his right-hand pocket. Each time he needs a match, he is equallylikely to take it from either pocket.

Consider the moment when the mathematician first discovers that one of his matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches, $k = 0, 1, \dots, N$, in the other box?

4.8 Example

- Solution Let E denote the event that the mathematician first discovers that the right-hand matchbox is empty and that there are k matches in the left-hand box at the time. Now, this event will occur if and only if the $(N + 1)$ choice of the righthand matchbox is made at the $(N + 1 + N - k)$ trial. Hence, from Equation (8.2) (with $p = \frac{1}{2}$, $r = N + 1$, and $n = 2N - k + 1$), we see that

$$P(E) = \binom{2N - k}{N} \left(\frac{1}{2}\right)^{2N - k + 1}$$

Since there is an equal probability that it is the left-hand box that is first discovered to be empty and there are k matches in the right-hand box at that time, the desired result is

$$2P(E) = \binom{2N - k}{N} \left(\frac{1}{2}\right)^{2N - k}$$

4.8 k th Moment of Negative Binomial Random Variable

- The k th moment:

$$\begin{aligned} E[X^k] &= \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r} \\ &= \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \binom{n}{r} p^{r+1} (1-p)^{n-r} \\ &= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)} \\ &= \frac{r}{p} E[(Y-1)^{k-1}] \end{aligned}$$

where Y is a negative binomial random variable with parameters $r+1, p$.

4.8 Mean and Variance of Negative Binomial Random Variable

- Setting $k = 1$ in the preceding equation yields

$$E[X] = \frac{r}{p}$$

- Setting $k = 2$ in the equation for $E[X^k]$

$$E[X^2] = \frac{r}{p}E[Y - 1] = \frac{r}{p} \left(\frac{r+1}{p} - 1 \right)$$

- Therefore,

$$\begin{aligned}\text{Var}(X) &= \frac{r}{p} \left(\frac{r+1}{p} - 1 \right) - \left(\frac{r}{p} \right)^2 \\ &= \frac{r(1-p)}{p^2}\end{aligned}$$

- You can also use $X = Y_1 + Y_2 + \cdots + Y_r$ to calculate the mean and variance.