Discrete Random Variables

Henry Pfister

ECE 581: Random Signals and Noise

Lecture 3

Slides courtesy of Christ Richmond with slight modifications.

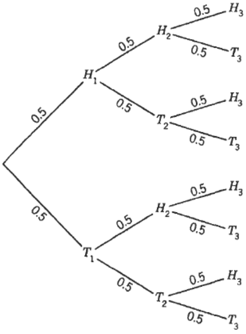
Table of contents

Remarks

- Until now, we've defined random experiments by identify/visualizing the sample space, and noting the assignment of probabilities.
  - Some experiments have non-numeric outcomes, e.g. a coin flip $S = \{H, T\}$.
  - Some experiments naturally have numeric outcomes, e.g. a die roll $S = \{1, 2, 3, 4, 5, 6\}$.
- Now, we consider mapping each experiment outcome to a numeric result. This enables rather insightful analysis via averages.
  - Various strategies exists to make this mapping/assignment.
  - The strategy adopted depends on ones needs/interest.

Random Variables

- A random variable is defined by a function (or mapping) which assigns a real number to each sample point in the sample space of an experiment.
- Consider 3 repeated fair coin flips:



Sample points	$P(\bullet)$	$K = \#\text{heads}$
$H_1 H_2 H_3$	0.125	3
$H_1 H_2 T_3$	0.125	2
$H_1 T_2 H_3$	0.125	2
$H_1 T_2 T_3$	0.125	1
$T_1 H_2 H_3$	0.125	2
$T_1 H_2 T_3$	0.125	1
$T_1 T_2 H_3$	0.125	1
$T_1 T_2 T_3$	0.125	0

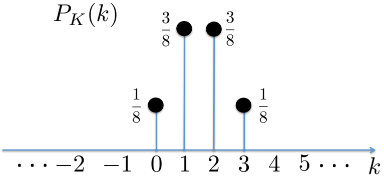
Clearly, outcomes for this sample space are non-numeric sequences, e.g. $H_1 T_2 T_3$.

Random Variables

- Appropriate random variables can be associated with this experiment. Consider the count of heads in each sequence. For example, let define $K = \#$ heads.
  - Do you recognize this process?
  - 3 coin flips are essentially repeated independent Bernoulli subexperiments. The random variable K has a Binomial distribution,

$$\Pr(K = k) = \begin{cases} \binom{3}{k} p^k (1-p)^{3-k}, & \text{when } k \in \{0,1,2,3\} \\ 0, & \text{when } k \notin \{0,1,2,3\}. \end{cases} \quad (1)$$

– Note that k has nonzero probability for discrete values of 0,1,2, and 3. Thus, we say k is a discrete random variable and {0,1,2,3} is called the range of k.
– We plot this for $p = 0.5$



- X is called a discrete random variable if its range is a countable set.[1]

[1] A set is countable if there is a one-to-one mapping between its members and the positive integers.

Probability Mass Function (PMF)

- Since a random variable is defined by assigning a real number to each outcome, each of these real numbers has a probability associated with it.
- The probability mass function (PMF) of discrete random variable X is function

$$P_X(x) = \Pr(X = x).$$

  - The function $P_X(x)$ is defined for all $x \in \mathbb{R}$
  - The value of $P_X(x)$ conveys the probability of event $\{X = x\}$.
  - Uppercase X indicates a random variable; lower case x represents a possible realization of X.
- Since the PMF defines a probability space, it also satisfies the axioms of probability.
- For a discrete random variable X with PMF $P_X(x)$ defined on range S_X:
  1. $P_X(x) \geq 0$ for all x.
  2. $\displaystyle\sum_{x \in S_X} P_X(x) = 1$.
  3. For any event $B \subset S_X$, it has probability $\Pr(B) = \displaystyle\sum_{x \in B} P_X(x)$.

Example 3.5 Problem

When the basketball player Wilt Chamberlain shot two free throws, each shot was equally likely either to be good (g) or bad (b). Each shot that was good was worth 1 point. What is the PMF of X, the number of points that he scored?

Quiz 3.2

The random variable N has PMF

$$P_N(n) = \begin{cases} c/n & n = 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Find

(a) The value of the constant c

(b) $P[N = 1]$

(c) $P[N \geq 2]$

(d) $P[N > 3]$

Families of Discrete Random Variables: Bernoulli

- We've seen various experiments that can be described by fundamental processes.
  - For example, a Bernoulli process can represents a coin flip, a pass/fail test, a win/lose choice, etc.
- Similarly, a set of basic random variables can represent various experiments.
  - You will notice that several of these random variable models we've already discussed as "processes."

Bernoulli Random Variable:

- X is a Bernoulli random variable if it has a PMF of form:

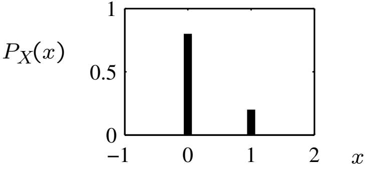
$$P_X(x) = \begin{cases} 1-p, & x=0 \\ p, & x=1 \\ 0, & \text{otherwise,} \end{cases} \quad \text{where } 0 \le p \le 1. \tag{2}$$

  At times we denote this by the shorthand notation $X \sim \text{Ber}(p)$.

- Several other r.v. types emerge as a result of repeated Bernoulli trials.

Example 3.8

If there is a 0.2 probability of a reject, the PMF of the Bernoulli (0.2) random variable is



$$P_X(x) = \begin{cases} 0.8 & x = 0, \\ 0.2 & x = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Families of Discrete R.V.'s: Binomial Random Variable

Binomial Random Variable:

- K is a binomial random variable if it has a PMF of form

$$P_K(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

  where $0 \leq p \leq 1$. This is often denoted as $K \sim \text{binomial}(n, p)$.

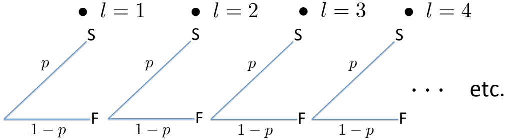
- Recall that the binomial distribution is obtained by counting the # ones (i.e. successes, heads, wins, etc.) occurring in a sequence of repeated Bernoulli trials.
  - The probability of obtaining k successes in n independent Bernoulli trials equals $P_K(k)$ in (3)
  - Recall examples 2.19, 2.20, and 2.21 Yates/Goodman: each can use binomial r.v. as a model.

- It should not be surprising that a binomial r.v. is obtained by adding together n independent Bernoulli random variables, i.e. $K = X_1 + X_2 + \cdots + X_n$ where $X_i \sim \text{Ber}(p)$, $i = 1, 2, \ldots, n$.

Families of Discrete R.V.'s: Geometric Random Variable

Geometric Random Variable:

- Consider repeating a Bernoulli experiment until the first success. For example, repeating a coin flip indefinitely until we obtain the first head.
- The sample space for this is illustrated below. Let the random variable L equal the number of trials before obtaining the first success.



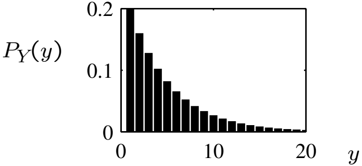
- Clearly, the illustrated sample space shows that the PMF for L is given by

$$P_L(l) = \begin{cases} p(1-p)^{l-1}, & l = 1, 2, 3, \ldots \\ 0, & \text{otherwise,} \end{cases} \quad \text{where } 0 \leq p \leq 1. \tag{4}$$

- We say $L \sim$ geometric(p) if it has a PMF of the form (4).

Example 3.10

If there is a 0.2 probability of a reject, the PMF of the geometric (0.2) random variable is



$$P_Y(y) = \begin{cases} (0.2)(0.8)^{y-1} & y = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

Families of Discrete R.V.'s: Uniform Random Variable

Uniform Random Variable:

- X is a discrete uniform random variable if, for $n \geq 1$, it has a PMF of the form

$$P_X(x) = \begin{cases} \frac{1}{n}, & \text{if } x \in \{k, k+1, k+2, \ldots, k+n-1\} \\ 0, & \text{otherwise.} \end{cases}$$

- Many experiments modeled via the uniform distribution: fair coin flips and die rolls

Poisson Random Variable:

- There are several problems in engineering and science where the arrival of specific phenomena are anticipated. For example,
  - the number of customers arriving in line at Starbucks during a fixed period of time
  - the number of order requests Amazon.com receives per minute
  - the number of particle emissions from a radio active source per second
  - the number of lightning strike incidents per hour world-wide
- It is useful to have a way of modeling the # of these events occuring per unit time.

Families of Discrete R.V.'s: Poisson Random Variable

- N is a Poisson (α) random variable if it has a PMF of form

$$P_N(n) = \begin{cases} \dfrac{\alpha^n e^{-\alpha}}{n!}, & n = 0, 1, 2, 3, \ldots; \text{ where } \alpha > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

- By convention, we refer to each counter increase as an "arrival."
- The Poisson PMF in (5) is often used to model the probability of observing N arrivals during some specified unit of time. This is captured by
  - assuming some arrival rate λ (average # arrivals per unit time) for the process of interest
  - and then specifying a duration of observation T;
  - PMF parameter is then determined as $\alpha = \lambda T$.

Example 3.17 Problem

The number of hits at a website in any time interval is a Poisson random variable. A particular site has on average $\lambda = 2$ hits per second. What is the probability that there are no hits in an interval of 0.25 seconds? What is the probability that there are no more than two hits in an interval of one second?

Cumulative Distribution Function (CDF)

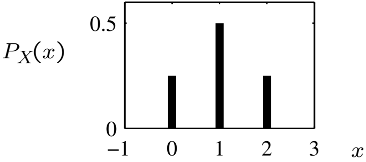
- The PMF of a random variable gives a complete picture of its probablistic behavior.
- There is another useful function that carries the same information.
- The cumulative distribution function (CDF) of random variable X is defined as

$$F_X(x) = \Pr[X \leq x] = \sum_{a \in S_X \,:\, a \leq x} P_X(a)$$

  where F_X is the typical notation for the CDF of X.
- The CDF also has a set of properties following from the axioms of probability
  1. $F_X(-\infty) = 0$ and $F_X(\infty) = 1$ [starts at zero ends at unity]
  2. $F_X(a) \leq F_X(b)$ for any $a \leq b$ in S_X [monotonically increasing function]
  3. $P_X(x) = F_X(x) - F_X(x - \epsilon)$ for enough small $\epsilon > 0$ [jumps given by PMF]
  4. $F_X(x) = F_X(a_i)$ for all $x \in S_X$ such that $a_i \leq x < a_{i+1}$ where $S_X = \{a_1, a_2, a_3, \ldots\}$.

Example 3.21 Problem

In Example 3.5, random variable X has PMF



$$P_X(x) = \begin{cases} 1/4 & x = 0, \\ 1/2 & x = 1, \\ 1/4 & x = 2, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

Find and sketch the CDF of random variable X.

CDF and Interval Probabilities

- For all $a \leq b$ in range S_X:

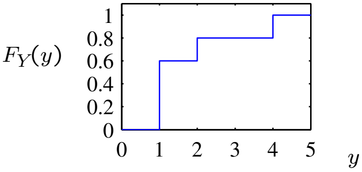
$$\Pr(a < X \leq b) = \sum_{\substack{x \in S_X : \\ a < x \leq b}} P_X(x)$$

$$= \sum_{\substack{x \in S_X : \\ x \leq b}} P_X(x) - \sum_{\substack{x \in S_X : \\ x \leq a}} P_X(x)$$

$$= F_X(b) - F_X(a).$$

Quiz 3.4

Use the CDF $F_Y(y)$ to find the following probabilities:



(a) $P[Y < 1]$
(b) $P[Y \le 1]$
(c) $P[Y > 2]$
(d) $P[Y \ge 2]$
(e) $P[Y = 1]$
(f) $P[Y = 3]$

Averages and Expected Values

- It is useful to define numbers (called statistics) that provide a quick summary of the behavior of a random variable or process.
- Statistics identifying points of central tendency include the mean, median, and mode.
- The mean or expected value or expectation of discrete r.v. X is

$$E[X] = \sum_{x \in S_X} x \cdot P_X(x).$$

  - The mean is the effective center of mass of the PMF.
- A median of a discrete r.v. X is a number x_{med} that satisfies

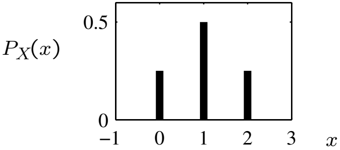
$$\Pr(X \leq x_{med}) \geq \frac{1}{2}, \text{ and } \Pr(X \geq x_{med}) \geq \frac{1}{2}.$$

- A mode of a discrete r.v. is a number x_{mod} that satisfies

$$P_X(x) \leq P_X(x_{mod}) \text{ for all } x \in S_X.$$

Example 3.24 Problem

Random variable X in Example 3.5 has PMF



$$P_X(x) = \begin{cases} 1/4 & x = 0, \\ 1/2 & x = 1, \\ 1/4 & x = 2, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

What is $E[X]$?

Find Mean, Mode and Median.

What about Binomial r.v. in (1)?

Averages and Expected Values

- Consider n independent observations $X_1, X_2, \ldots, X_n$ of a r.v. and its sample average:

$$s_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{x\in S_X} k(x)\cdot x = \sum_{x\in S_X}\frac{k(x)}{n}\cdot x$$

  where $k(x)$ is the number of occurences of outcome $X_i = x$, $i = 1, 2, \ldots, n$.

- As $n \to \infty$ we expect that $\frac{k(x)}{n} \to P_X(x)$ by the relative frequency definition of probability:

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty}\sum_{x\in S_X}\frac{k(x)}{n}\cdot x = \sum_{x\in S_X}\lim_{n\to\infty}\frac{k(x)}{n}\cdot x = \sum_{x\in S_X} P_X(x)\cdot x = E[X].$$

  - Later, we will investigate the conditions under which this is always true.
  - Intuitively, $E[X]$ equals the average a large number of independent realizations $\frac{1}{n}\sum_{i=1}^{n} X_i$

Example Calculations of Expected Values (1)

- A Bernoulli(p) r.v. X has mean $E[X] = p$:

$$E[X] = \sum_{x \in S_X} x \cdot P_X(x) = 0 \cdot (1-p) + 1 \cdot p = p.$$

- Geometric (p) r.v. L has mean $E[L] = 1/p$:

$$\begin{aligned} E[L] &= \sum_{l \in S_L} l \cdot P_L(l) = \sum_{l=1}^{\infty} l \cdot p \cdot (1-p)^{l-1} = p \sum_{l=1}^{\infty} -\frac{d}{dp}(1-p)^l \\ &= -p \frac{d}{dp} \sum_{l=1}^{\infty} (1-p)^l = -p \frac{d}{dp} \left(-1 + \sum_{l=0}^{\infty} (1-p)^l \right) \\ &= -p \frac{d}{dp} \left(-1 + \frac{1}{1-(1-p)} \right) = -p \frac{d}{dp} \left(-1 + \frac{1}{p} \right) = \frac{1}{p}. \end{aligned}$$

where third equality follows from well-known derivate $\frac{d}{dy} y^n = n y^{n-1}$.

Example Calculations of Expected Values (2)

- Poisson (α) r.v. N has mean $E[N] = \alpha$:

$$E[N] = \sum_{n \in S_N} n \cdot P_N(n) = \sum_{n=0}^{\infty} n \cdot \frac{\alpha^n e^{-\alpha}}{n!} = \sum_{n=1}^{\infty} \alpha e^{-\alpha} \cdot \frac{\alpha^{n-1}}{(n-1)!}$$
$$= \alpha e^{-\alpha} \sum_{n=1}^{\infty} \frac{\alpha^{n-1}}{(n-1)!} = \alpha e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = \alpha e^{-\alpha} \cdot e^{\alpha} = \alpha$$

  where the third to last equality follows from variable change $k = n - 1$; second to last equality follows from exponent Taylor series, i.e. $e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}$.

- Recall $\alpha = \lambda T$; thus, since $E[N] = \alpha$ we have $\lambda = E[N]/T$;
  - justifies λ as arrival rate, i.e. the average number of arrivals per unit time.

Example Calculations of Expected Values (3)

- The Binomial (n, p) r.v. has mean $E[K] = np$ (to be proven later).
- The Uniform r.v. has mean $E[X] = k + \frac{(n-1)}{2}$, i.e. the midpoint:

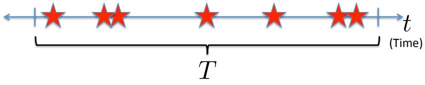
$$E[X] = \sum_{x \in S_X} x \cdot P_X(x) = \sum_{x=k}^{k+n-1} x \cdot \frac{1}{n} = \frac{1}{n}[k + (k+1) + (k+2) + \cdots + (k+n-1)]$$

$$= \frac{1}{n}[n \cdot k + (0 + 1 + 2 + 3 + \cdots + n - 1)] = k + \frac{1}{n}\left[\frac{(n-1)n}{2}\right]$$

$$= k + \frac{(n-1)}{2} = \frac{k + (k+n-1)}{2}.$$

Note on Poisson Process

- The arrivals of a Poisson process can be seen as instances on a continuous time axis:



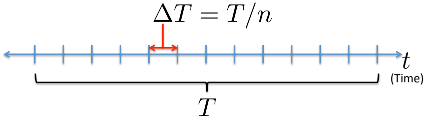
  - For an interval of duration T, the Poisson distribution models the probability of receiving $k \in \{0, 1, 2, \ldots\}$ arrivals.

- Recall that the Binomial distribution describes the probability of receiving $k \in \{0, 1, \ldots, n\}$ successes in n independent Bernoulli trials.

- It will now be shown that a Poisson distribution is the limiting distribution of a judiciously chosen Binomial processs.

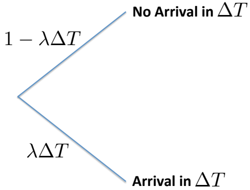
Understanding Poisson Processes (1)

- Consider dividing an interval of duration T into n equal slots of duration $\Delta T = T/n$, where n is large enough that we assume only one arrival is possible during a slot:



  - Since 0 or 1 arrivals occur in each slot, the arrivals are modeled as a Bernoulli (p) process.
    - The expected number of successes in n independent repetitions of Bernoulli trials (i.e. expected value of a Binomial (n, p) r.v.) is given by np.
    - Recall expected value of a Poisson distribution is $\alpha = \lambda T$.
    - Thus, matching expectations means that $np = \alpha \implies p = \alpha/n = \lambda n \Delta T/n = \lambda \Delta T$.

–Using a probability of success given by $p = \lambda \Delta T = \lambda T/n$ we have Bernoulli trial:



Understanding Poisson Processes (2)

- Consider a Binomial $(n, p = \alpha/n = \lambda T/n)$ r.v. with PMF

$$P_{K_n}(k) = \binom{n}{k}\left(\frac{\alpha}{n}\right)^k\left(1-\frac{\alpha}{n}\right)^{n-k}, \quad k = 0, 1, 2, \ldots, n-1, n. \tag{6}$$

- This Binomial distribution characterizes the count of # arrivals in duration T:
  - Each time slot is an iid Bernoulli trial, with success probability $p = \lambda\Delta T = \lambda T/n$.
  - The Bernoulli trial for any time slot is independent of the Bernoulli trials in other time slots. Thus, nonoverlapping time intervals are mutually independent and the process has no memory.
  - Knowing no arrivals occurred for an hour does not affect chance of an arrival in the next 5 mins.

- The limiting behavior of (6) as $n \to \infty$ (or $\Delta T \to 0$) is given by

$$\binom{n}{k}\cdot\frac{1}{n^k} = \frac{n!}{k!\cdot(n-k)!\cdot n^k} = \frac{n(n-1)(n-2)\cdots(n-k)(n-k-1)\cdots 2\cdot 1}{(n-k)(n-k-1)\cdots 2\cdot 1\cdot k!\cdot n^k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!\cdot n^k} = \frac{n^k + a_{k-1}n^{k-1} + \cdots + a_1 n + a_o}{k!\cdot n^k}$$

$$= \frac{1 + a_{k-1}\frac{n^{k-1}}{n^k} + \cdots + a_1\frac{n}{n^k} + a_o\frac{1}{n^k}}{k!} \to \frac{1}{k!} \text{ as } n \to \infty$$

Understanding Poisson Processes (3)

- Further note that for fixed k we have limits[1]

$$\lim_{n\to\infty}\left(1-\frac{\alpha}{n}\right)^{n}=e^{-\alpha},\quad \text{and} \quad \lim_{n\to\infty}\left(1-\frac{\alpha}{n}\right)^{-k}=1.$$

Thus, we have

$$\lim_{n\to\infty} P_{K_n}(k)=\frac{1}{k!}\cdot\alpha^{k}\cdot e^{-\alpha},\quad k=0,1,2,3,\ldots$$

that we recognize as the Poisson distribution.

[1] Recall from calculus that $\lim_{s\to\infty}\left(1+\frac{q}{s}\right)^{s}=e^{q}$.

Quiz 3.5

In a pay-as-you go cellphone plan, the cost of sending an SMS text message is 10 cents and the cost of receiving a text is 5 cents. For a certain subscriber, the probability of sending a text is 1/3 and the probability of receiving a text is 2/3. Let C equal the cost (in cents) of one text message and find

(a) The PMF $P_C(c)$

(b) The expected value $E[C]$

(c) The probability that the subscriber receives four texts before sending a text.

(d) The expected number of texts received by the subscriber before the subscriber sends a text.

Functions of a Random Variable (1)

- There are many situations where the random values of some experiment outcome are used to compute other quantities.
  - For example, let the number of open cash register lines in a supermarket be C. This number may be determined by the number of customers K arriving per minute, such that[2] $C = \lceil \frac{1}{5} K \rceil$.
  - Given two rolls of a four-sided die yielding (D_1, D_2), $D_i \in \{1, 2, 3, 4\}$, $i = 1, 2$; perhaps we're interested in the larger of the two numbers, i.e. $X = \max(D_1, D_2)$.
  - Given r.v.'s $X_1, X_2, \ldots, X_n$, the sample mean is $\frac{1}{n} \sum_{k=1}^{n} X_i$.
- Let $g(x)$ be some arbitrary function defined on the range of X, i.e. for $x \in S_X$. If $Y = g(X)$ where X is a r.v., the Y is also a r.v.; we refer to Y as a derived random variable.
- A function of random variables is itself a random variable. Thus, it is also described by some PMF, i.e. there is some $P_Y(y)$ defined for $y \in S_Y$.

[2] The "ceiling" of x, denoted $\lceil x \rceil$, is the smallest integer greater than or equal to x. Similarly, the "floor" of x, denoted $\lfloor x \rfloor$, is the greatest integer less than or equal to x.

Functions of a Random Variable (2)

- **Theorem:** Given discrete r.v. X with PMF $P_X(x)$, the PMF of derived r.v. $Y = g(X)$ is

$$P_Y(y) = \Pr(Y = y) = \Pr[g(X) = y] = \sum_{x \in S_X : g(x) = y} P_X(x).$$

- Note that if function $g(x)$ is one-to-one (i.e. $y = g(x) \implies x = g^{-1}(y)$), then

$$P_Y(y) = \Pr(Y = y) = \Pr[g(X) = y] = \Pr[X = g^{-1}(y) = x] = P_X(x).$$

Quiz 3.6

Monitor three customers purchasing smartphones at the Phonesmart store and observe whether each buys an Apricot phone for \$450 or a Banana phone for \$300. The random variable N is the number of customers purchasing an Apricot phone. Assume N has PMF

$$P_N(n) = \begin{cases} 0.4 & n = 0, \\ 0.2 & n = 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

M dollars is the amount of money paid by three customers.

(a) Express M as a function of N.

(b) Find $P_M(m)$ and $\mathrm{E}[M]$.

Expected Value of a Derived Random Variable

- Given discrete r.v. $X \sim P_X(x)$, the expected value of derived r.v. $Y = g(X)$ is

$$E[Y] = \sum_{y \in S_Y} y \cdot P_Y(y) = \sum_{y \in S_Y} y \sum_{x \in S_X : g(x)=y} P_X(x)$$
$$= \sum_{y \in S_Y} \sum_{x \in S_X : g(x)=y} g(x) \cdot P_X(x) = \sum_{x \in S_X} g(x) \cdot P_X(x) = E[g(X)].$$

  - Thus, expected value of derived r.v. $Y = g(X)$ can be found without knowing PMF $P_Y(y)$.

- Given discrete r.v. X with PMF $P_X(x)$, consider the derived r.v. $Y = aX + b$, $a, b \in \mathbb{R}$:

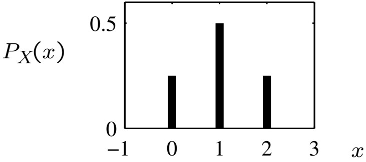
$$E[Y] = E[aX + b] = \sum_{x \in S_X} (ax + b) \cdot P_X(x) = \sum_{x \in S_X} [ax \cdot P_X(x) + b \cdot P_X(x)]$$
$$= a \sum_{x \in S_X} x \cdot P_X(x) + b \sum_{x \in S_X} P_X(x) = a \cdot E[X] + b \cdot 1.$$

  Thus, a linear transformation of a r.v. has mean $E[aX + b] = aE[X] + b$.

- What is the expected value of $Y = X - E[X]$?

Example 3.30 Problem

Recall from Examples 3.5 and 3.24 that X has PMF



$$P_X(x) = \begin{cases} 1/4 & x = 0, \\ 1/2 & x = 1, \\ 1/4 & x = 2, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

What is the expected value of $V = g(X) = 4X + 7$?

Example 3.29 Problem

In Example 3.26,

$$P_X(x) = \begin{cases} 1/4 & x = 1, 2, 3, 4, \\ 0 & \text{otherwise}, \end{cases}$$

$$Y = g(X) = \begin{cases} 105X - 5X^2 & 1 \le X \le 5, \\ 500 & 6 \le X \le 10. \end{cases}$$

What is E$[Y]$?

Expected Value of a Derived Random Variable

- Given discrete r.v. $X \sim P_X(x)$, consider r.v. $Y = g(X) = a_1 g_1(X) + a_2 g_2(X)$, where $a_1, a_2 \in \mathbb{R}$:

$$E[Y] = E[g(X)] = \sum_{x \in S_X} g(x) \cdot P_X(x) = \sum_{x \in S_X} [a_1 g_1(x) + a_2 g_2(x)] \cdot P_X(x)$$
$$= a_1 \sum_{x \in S_X} g_1(x) \cdot P_X(x) + a_2 \sum_{x \in S_X} g_2(x) \cdot P_X(x) = a_1 \cdot E[g_1(X)] + a_2 E[g_2(X)].$$

- Thus, a linear combination of functions has mean

$$E[a_1 g_1(X) + a_2 g_2(X)] = a_1 E[g_1(X)] + a_2 E[g_2(X)].$$

- Key idea: Expectation is a linear operation

Variance, Standard Deviation, and Moments

- The PMF completely describes a r.v.; we can deduce many things from PMF $P_X(x)$.
- Recall that expected value or mean of a r.v. provides a measure of central tendancy.
  - $E[X] = \sum_{x \in S_X} x \cdot P_X(x)$ is the center of mass of PMF.
  - $E[X]$ is only a single number (summarizes only one aspect of r.v.)
- How much might does X deviates from $E[X]$? Deviation can be quantified in many ways:
  - absolute distance $|X - E[X]|$
  - squared distance $(X - E[X])^2$
  - or other valid distance metric $d(X, E[X])$, etc.
- Squared distance $(X - E[X])^2$ is popular choice due to ease of use.
- Average squared distance from distribution mean $E[X]$ is the variance of r.v. X:

$$\mathrm{Var}[X] = E\{(X - E[X])^2\} = \sum_{x \in S_X} (x - E[X])^2 \cdot P_X(x).$$

- For a r.v. X, typical notation is $\mu_X \triangleq E[X]$ and $\sigma_X^2 \triangleq \mathrm{Var}[X]$.

Variance, Standard Deviation, and Moments

- The variance can be written

$$\begin{aligned}\text{Var}[X] &= E\left[(X - E[X])^2\right] \\ &= E\left[X^2 - 2X \cdot E[X] + E[X]^2\right] \\ &= E[X^2] - 2E[X] \cdot E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2.\end{aligned}$$

- Since $(X - E[X])^2 \geq 0$, it follows that

$$\text{Var}[X] = \sum_{x \in S_X} (x - E[X])^2 \cdot P_X(x) \geq 0, \quad \text{always}.$$

  - Note that if $\text{Var}[X] = 0$, then it follows that $X = E[X]$ deterministically.

- The standard deviation of r.v. X is the square root of the variance:

$$\sigma_X = \sqrt{\text{Var}[X]} = \sqrt{E[X^2] - E^2[X]}.$$

  - Note σ_X and $E[X]$ have the same units as X.

Guessing the Value of X as the Mean, i.e. $\widehat{X} = E[X]$

- How should we guess the value that r.v. X will assume?...call it $\widehat{X}$.
- Consider choosing $\widehat{X}$ to minimize $E\left[(\widehat{X}-X)^2\right]$, i.e. the mean squared error (MSE):

$$E\left[(\widehat{X}-X)^2\right] = \sum_{x\in S_X} (\widehat{X}-x)^2 P_X(x)$$

$$\frac{d}{d\widehat{X}} E\left[(\widehat{X}-X)^2\right] = \frac{d}{d\widehat{X}} \sum_{x\in S_X} (\widehat{X}-x)^2 P_X(x) = \sum_{x\in S_X} 2(\widehat{X}-x) P_X(x)$$

- Solving for the unique stationary point gives

$$0 = 2\widehat{X} \sum_{x\in S_X} P_X(x) - 2 \sum_{x\in S_X} x \cdot P_X(x) = 2\widehat{X} - 2E[X]$$

$$\implies \widehat{X} = E[X].$$

Thus, $\widehat{X} = E[X]$ is the guess that minimizes the MSE.

- Since $E\left[(\widehat{X}-X)^2\right] = E\left[(X-\widehat{X})^2\right]$, the minimum MSE is $E\left[(X-E[X])^2\right]$, i.e. the variance $\mathrm{Var}[X]$.

Moments of a R.V.

- The *n*-th moment of r.v. $X \sim P_X$ is

$$E[X^n] = \sum_{x \in S_X} x^n \cdot P_X(x).$$

- The *n*-th central moment of r.v. $X \sim P_X$ is

$$E\{(X - E[X])^n\} = \sum_{x \in S_X} (x - E[X])^n \cdot P_X(x).$$

  - The mean $E[X]$ is the first moment and $E[X^2]$ is the second moment;
  - The variance $\sigma_X^2 = E\{(X - E[X])^2\}$ is the second central moment.

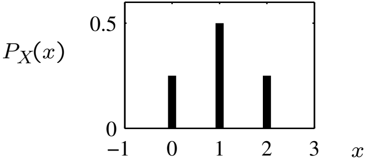
- Consider derived r.v. $Y = aX + b$. What's it's variance $\text{Var}[Y] = E[Y^2] - E^2[Y] = ?$

$$\begin{aligned} E[Y^2] = E[(aX+b)^2] = E[a^2X^2 + 2abX + b^2] &= a^2E[X^2] + 2abE[X] + b^2 \\ E[Y]^2 = (aE[X] + b)^2 &= a^2E[X]^2 + 2abE[X] + b^2 \\ E[Y^2] - E^2[Y] &= a^2(E[X^2] - E[X]^2) = a^2\sigma_X^2 \end{aligned}$$

  Thus, $\text{Var}[aX + b] = a^2\text{Var}[X]$.

Example 3.32 Problem

Continuing Examples 3.5, 3.24, and 3.30, we recall that X has PMF



$$P_X(x) = \begin{cases} 1/4 & x = 0, \\ 1/2 & x = 1, \\ 1/4 & x = 2, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

and expected value $E[X] = 1$. What is the variance of X?

Example 3.33 Problem

A printer automatically prints an initial cover page that precedes the regular printing of an X page document. Using this printer, the number of printed pages is $Y = X + 1$. Express the expected value and variance of Y as functions of E$[X]$ and Var$[X]$.

Determine expected value and variance of $Y = (X - E[X])/\sigma_X$, i.e. *standardized* r.v.

Quiz 3.8

In an experiment with three customers entering the Phonesmart store, the observation is N, the number of phones purchased. The PMF of N is

$$P_N(n) = \begin{cases} (4-n)/10 & n = 0,1,2,3 \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Find

(a) The expected value $E[N]$

(b) The second moment $E[N^2]$

(c) The variance $\text{Var}[N]$

(d) The standard deviation σ_N

More on Variance and Some Examples

- $X \sim$ Bernoulli (p):

$$E[X] = 0 \cdot (1-p) + 1 \cdot p = p, \quad E[X^2] = 0^2 \cdot (1-p) + 1^2 \cdot p = p,$$
$$\text{Var}[X] = E[X^2] - E^2[X] = p - p^2 = p(1-p).$$

- $N \sim$ Poisson (α): (note that: $E[N(N-1)] = E[N^2] - E[N]$)

$$E[N(N-1)] = \sum_{n=0}^{\infty} n(n-1) P_N(n) = \sum_{n=0}^{\infty} n(n-1) \frac{\alpha^n e^{-\alpha}}{n!} = \sum_{n=2}^{\infty} n(n-1) \frac{\alpha^n e^{-\alpha}}{n!}$$
$$= \alpha^2 e^{-\alpha} \sum_{n=2}^{\infty} \frac{\alpha^{n-2}}{(n-2)!} = \alpha^2 e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = \alpha^2 e^{-\alpha} e^{\alpha} = \alpha^2.$$

Thus, $E[N^2] = \alpha^2 + E[N] = \alpha^2 + \alpha$ and $\text{Var}[N] = E[N^2] - E[N]^2 = \alpha^2 + \alpha - \alpha^2 = \alpha.$

More on Variance and Some Examples

- $L \sim$ Geometric (p):

$$\text{Setup: } E[L] = \frac{1}{p}, \quad E[L^2] = \sum_{l \in S_L} l^2 \cdot P_L(l) = \sum_{l=1}^{\infty} l^2 \cdot p \cdot (1-p)^{l-1}.$$

$$\text{Note: } (1-p)\frac{d^2}{dp^2}(1-p)^l = l^2(1-p)^{l-1} - l(1-p)^{l-1}$$

$$E[L^2] = p\sum_{l=1}^{\infty} l^2(1-p)^{l-1} = p\sum_{l=1}^{\infty}(1-p)\frac{d^2}{dp^2}(1-p)^l + l(1-p)^{l-1}$$

$$= p(1-p)\frac{d^2}{dp^2}\sum_{l=1}^{\infty}(1-p)^l + E[L] = p(1-p)\frac{d^2}{dp^2}\left(-1+\frac{1}{p}\right) + E[L]$$

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

$$\implies \text{Var}[L] = E[L^2] - E[L]^2 = \frac{1-p}{p^2}.$$

More on Variance and Some Examples

- $X \sim$ Uniform on $\{k, k+1, \ldots, k+n-1\}$[3]
- Assume $k = 1$ and check that $E[X] = \frac{n+1}{2}$

$$E[X^2] = \sum_{x=1}^{n} x^2 \cdot \frac{1}{n} = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6} \implies$$

$$\mathrm{Var}[X] = E[X^2] - E^2[X] = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= (n+1)\left[\frac{2(2n+1) - 3(n+1)}{12}\right] = (n+1)\frac{n-1}{12} = \frac{n^2-1}{12}$$

Since variance is not changed by translation/shift, this is also variance for $k \neq 1$

[3] Can verify sum of squares given by $\sum_{k=1}^{n} k^2 = 1 + 2^2 + 3^2 + \cdots + (n-1)^2 + n^2 = \frac{n(n+1)(2n+1)}{6}$.

References

[1] Yates and Goodman, *Probability and Stochastic Processes*, 3rd Ed., Wiley, 2014.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, Athena Sci., 2002.

[3] A. W. Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill Inc., 1967.