EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

Lecture 5 - Outline

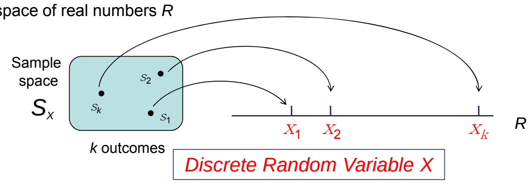
1. Random Variables
2. Discrete Random Variables.

Chapter 2 Discrete Random Variables

Random Variable (RV)

An assignment of a value (real number) to each & every outcome in a probability space

Mathematically: A random variable X is a function from the sample space S_X to the space of real numbers R



Notations:

The random variable is denoted by capital letter: X

& the values it takes are denoted by lower-case italic: x

Types of Random Variables

- Discrete random variable

  Possible values (range) of X is a countable set $\{x_1, x_2, \ldots\}$
- Finite random variable

  Range of X is a finite set $\{x_1, x_2, \ldots, x_k\}$
- Continuous random variable

  Range of X is neither countable nor finite.

Odd concept in Book. Improper Experiment: Random variable not defined in all of outcome space; some outcomes lead to undefined random variables
Common definition: random variable defined for all outcomes.

Examples of Discrete Random Variables

- Number of Geiger-counter counts n in 1 sec (n = 0,1,2,..)
- Number of photons n collected by a photodetector in 1 μsec (n = 0,1,2,..)
- Fabricated semiconductor chip is tested for faults. Random variable X assigns 0 if chip is faulty, 1 if OK. X is discrete and finite.

2.1 Probability Mass Function (PMF)

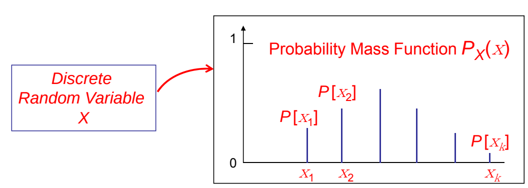
A *random variable* X that takes values $\{x_1, x_2, \ldots, x_k\}$

associated with outcomes $\{s_1, s_2, \ldots, s_k\}$

is characterized by the probabilities

$\{P[X=x_1] = P[s_1], P[X=x_2] = P[s_2], \ldots, P[X=x_k] = P[s_k]\}$

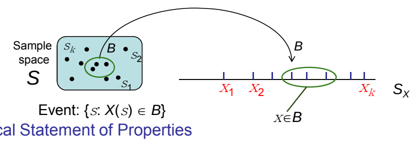
These probabilities are also written as $P_X(x)$ or simply $P[x]$



Note: We can use PMFs directly to compute probabilities without going back to original outcomes in probability space…

Properties of PMF

- $P_X(x_i) \geq 0, \quad i = 1, 2, \cdots k$
- $\sum_i P_X(x_i) = 1$



Event: $\{s: X(s) \in B\}$

Mathematical Statement of Properties

– For any x in S_X, $P_X(x) \geq 0$

– $\sum_{x \text{ in } S_X} P_X(x) = 1$

– For any set $B \subset S_X$, the probability that X is in B is

$$P[\{s \in S: X(s) \in B\}] \equiv P[B] = \sum_{x \text{ in } B} P_X(x)$$

Probability in original experiment is mapped into probability on subsets of the real line!

These properties follows from the underlying definition of probability space

Example 1

X = number of coin tosses until first head

– assume independent tosses, $P[H] = p > 0$

– $P_X(k) = P[X = k] = P[TT \cdots TH]$

– $P_X(k) = (1 - p)^{k-1}p, \quad k = 1, 2, \ldots$

Example 2

Random variable N has PMF $P_N(n) = c/n$, $n = 1, 2, 3, 4$, or 0 otherwise

–Find value of c: $c(1 + 1/2 + 1/3 + 1/4) = 1 \rightarrow c = 12/25$

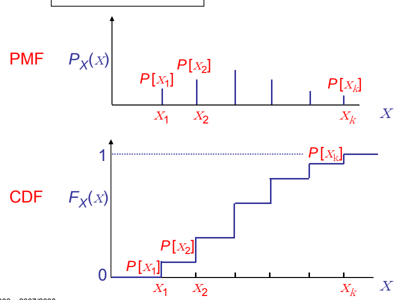
–Find $P[N \geq 2]$: $c(1/2 + 1/3 + 1/4)$

–Find $P[N < 3]$: $c(1 + 1/2)$

Cumulative Distribution Function (CDF)

The CDF of a random variable X is the total probability mass from $-\infty$ up to (and including) the point x.

$$F_X(x) = P[X \leq x] = P[X \in (-\infty, x]], \quad -\infty < x < \infty$$



Properties

- The CDF is a non-negative real-valued function $F_X(x) \in [0,1]$ defined for all real values of its argument x
- The CDF of any discrete random variable is a staircase function. If X takes on values $x_1, x_2, \ldots x_k$, with probabilities $P[x_1], P[x_2], \ldots, P[x_k]$, then the CDF has jumps at $x_1, x_2, \ldots x_k$, with heights $P[x_1], P[x_2], \ldots, P[x_k]$ and is flat in between the jumps

– $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$

– $F_X(\infty) = \lim_{x \to \infty} F_X(x) = 1$

– $F_X(x)$ is monotone non – decreasing in x

– $F_X(x)$ is piecewise constant and jumps at points $x \in S_X$ such that $P_X(x) > 0$

– For all $b \geq a, F_X(b) - F_x(a) = P[a < X \leq b]$

CDF is not a very useful for Discrete RVs

PMF is easier to use, has all information needed. So, why bother?

Because concept extends to ALL random variables, continuous, discrete, etc.

2.1 Statistics of Random Variables

Sample Statistics

A statistic based on a collection of sample values of a random variable is a single real number such as the average value, the median, or the standard deviation of the sample values.

Example

Scores in midterm

X = score of a randomly picked student (uniformly)

$P_X(x)$ = fraction of students with grade x

$$\text{Mean } m_X = \sum_{x=0}^{100} x P_X(x)$$

$$\text{Variance } = \sum_{x=0}^{100} (x - m_X)^2 P_X(x)$$

$$\text{Standard Deviation } \sigma_X = \sqrt{\text{Var[X]}}$$

Random Variable Statistics

A statistic of a RV maps the RV into a real-valued quantity (computed from the PMF).

Expected Value

The expected value ("average" or "mean") of a random variable X is

$$E[X] = \sum_{x \in S_X} x p_X(x) \quad \text{also denoted } m_X = E[X]$$

Interpretations:

– It is the weighted average of all possible values, using the PMF weights

– Center of gravity of PMF

– Average in large number of repetitions of the experiment (to be substantiated later in this course)

Variance and Standard Deviation

$$\text{Var}[X] = E[(X - m_X)^2] = \sum_{x \in S_x} (x - m_x)^2 P_X(x)$$

Standard deviation $\sigma_X = \sqrt{\text{Var}[X]}$

Moments $E[X^n] = n^{th}$ moment

Central Moments $E[(X - m_X)^n] = n^th$ central moment

Variance = Second central moment = Second moment - Mean^2

Proof

$$\begin{aligned}\text{Var}[X] &= \sum_{x \in S_X} (x - m_X)^2 P_X(x) \\ &= \sum_{x \in S_X} (x^2 - 2xm_X + m_X^2) P_X(x) \\ &= \sum_{x \in S_X} x^2 P_X(x) - 2m_X \sum_{x \in S_X} x P_X(x) + m_X^2 \sum_{x \in S_X} P_X(x) \\ &= E[X^2] - 2m_X m_X + m_X^2 = E[X^2] - m_X^2\end{aligned}$$

Other Simple Statistics of Random Variable X

Median = any number x_{med} such that

$P[X < x_{\text{med}}] = P[X > x_{\text{med}}]$

May not exist

May not be unique

Mode = any number x_{mod} such that

$P[X = x_{\text{mod}}] \geq P[X = x]$ for all $x \in S_X$

May not be unique, but must exist

2.2 Families of Discrete Random Variables

1) Uniform
2) Bernoulli
3) Geometric
4) Binomial
5) Poisson

These families of discrete RVs describe common experiments.

Members of each family differ only in the value of some experimental parameters.

Sequences and Series (p. 508)

$$B.4 \quad \sum_{i=0}^{n} q^i = 1 + q + \ldots + q^n = \frac{1 - q^{n+1}}{1 - q}$$

$$B.5 \quad \sum_{i=0}^{\infty} q^i = \frac{1}{1 - q}, |q| < 1$$

$$B.6 \quad \sum_{i=1}^{n} i q^i = \frac{q(1 - q^n[1 + n(1 - q)])}{(1 - q)^2}$$

$$B.7 \quad \sum_{i=1}^{\infty} i q^i = \frac{q}{(1 - q)^2}, |q| < 1$$

$$B.8 \quad \sum_{j=1}^{n} j = \frac{n(n+1)}{2}$$

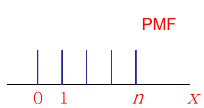
$$B.9 \quad \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{k=0}^{\infty} \frac{a^k}{k!} = e^a \text{ for all } a$$

1) Uniform Random Variables

Range: {0, 1, 2, …, n}

PMF: $P_X(x) = 1 / (n + 1)$ parameter n



Mean: $n / 2$

Variance: $n(n + 2) / 12$

Application: Used to model random experiments where all outcomes are equally likely, e.g., fair coin tosses, die rolls, roulette wheels, etc

Proof

$$E[X] = \sum_{x=0}^{n} p_X(x) x = \sum_{x=0}^{n} \frac{1}{n+1} x = \frac{1}{n+1} \sum_{x=0}^{n} x = \frac{1}{n+1} \frac{n(n+1)}{2} = \frac{n}{2}$$

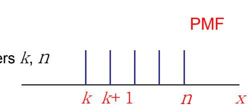
$$E[X^2] = \sum_{x=0}^{n} p_X(x) x^2 = \frac{1}{n+1} \sum_{x=0}^{n} x^2 = \frac{1}{n+1} \frac{n(n+1)(2n+1)}{6} = \frac{n(2n+1)}{6}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{n(2n+1)}{6} - (\frac{n}{2})^2 = \frac{n(n+2)}{12}$$

More General Uniform Random Variables

Range: {k, k+1, …, n}

PMF: $P_X(x) = 1/(n - k + 1)$ parameters k, n



Mean: $(n + k) / 2$

Variance: (n-k)(n-k+2)/12 (same for Uniform {0,…,n-k})

Proof

$$\begin{aligned}E[X] &= \sum_{j=k}^{n} j \frac{1}{n - k + 1} = \frac{1}{n - k + 1} \sum_{j=1}^{n-k+1} (k - 1 + j) \\ &= \frac{1}{n - k + 1}[(n - k + 1)(k - 1) + \frac{(n - k + 1)(n - k + 2)}{2}] \\ &= \frac{n - k + 2 + 2k - 2}{2} = \frac{n + k}{2}\end{aligned}$$

2) Bernoulli Random Variables

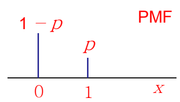
Range: {0, 1} (={failure, success}) Binary RV

PMF: $P_X(1) = p$, $P_X(0) = 1 - p$, parameter $p \in (0,1)$

mean: p

Variance: $p(1 - p)$

Application: Failure models, Bit errors



Proof

$$E[X] = \sum_{k=0}^{1} kP_X(k) = 0 \cdot (1-p) + 1 \cdot p = p$$
$$E[X^2] = \sum_{k=0}^{1} k^2 P_X(k) = 0 \cdot (1-p) + 1 \cdot p = p$$
$$\text{Var}[X] = p - p^2 = p(1-p)$$

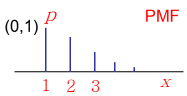
3) Geometric Random Variables

Range: {1, 2, … }

PMF: $P_X(x) = p(1 - p)^{x-1}$ parameter $p \in (0,1)$

Mean= $1 / p$

Variance= $(1 - p) / p^2$



Application:

- Waiting time to first "success" for repeated independent Bernoulli experiments
- Each bit in a communications channel has probability of error p; PMF of first bit that is in error is geometric
- Similar examples using sequential testing of potentially faulty components

Proof

$$E[X] = \sum_{k=1}^{\infty} kP_X(k) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$
$$= p\sum_{k=1}^{\infty} -\frac{d}{dp}(1-p)^k = p(-\frac{d}{dp})\sum_{k=1}^{\infty}(1-p)^k$$
$$= p(-\frac{d}{dp}\frac{(1-p)}{p}) = p\frac{1-p+p}{p^2} = \frac{1}{p}$$
$$E[X^2] = \sum_{n=1}^{\infty} n^2 p(1-p)^{n-1} = p(1-p)\sum_{n=1}^{\infty} n^2(1-p)^{n-2}$$
$$= p(1-p)\sum_{n=1}^{\infty} n(n-1)(1-p)^{n-2} + p\sum_{n=1}^{\infty} n(1-p)^{n-1}$$
$$= p(1-p)\frac{d^2}{dp^2}\frac{(1-p)}{p} - p\frac{d}{dp}\frac{(1-p)}{p}$$
$$= \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$
$$\text{Var}(X) = E[X^2] - E[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

4) Binomial Random Variable

Range: {0, 1, .., n}

PMF: $P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ parameters $p \in (0,1)$ and n

Mean: np

Variance: $np(1 - p)$

Application: Number of failures in n independent binary experiments

Proof

$$E[X] = \sum_{k=0}^{n} kp_X(k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$$
$$= np \text{ (sum of n independent Bernoulli variables)}$$

Example

- Service facility design (e.g. tellers at a bank)
  - n: number of customers
  - p: prob. customer requires service
  - s: no. of service persons
  - X: no. of service requests (RV)
- What is probability that customer has to wait?

$$P[X > s] = P[X = s+1] + \ldots + P[X = n]$$
$$= \sum_{j=s+1}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$

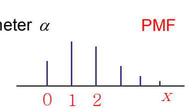
5) Poisson Random Variables

Range: {0, 1, 2, … }

PMF: $P_X(x) = \frac{\alpha^x}{x!}e^{-\alpha}$, $x = 0,1,\ldots$ parameter α

Mean: α

Variance: α



Applications:

Models the random number of events X ("arrivals") arriving within a fixed time window (e.g., photons or packets arriving at a communications switch or computer node).

Arrival rate: λ per time unit

Time interval over which arrivals occur: T

$$\alpha = \lambda T$$

Proof

$$E[X] = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=0}^{\infty} k \frac{\alpha^k}{k!} e^{-\alpha}$$
$$= e^{-\alpha} \sum_{k=0}^{\infty} k \frac{\alpha^k}{k!} = e^{-\alpha} \sum_{k=1}^{\infty} \frac{\alpha^k}{(k-1)!}$$
$$= \alpha e^{-\alpha} \sum_{k=1}^{\infty} \frac{\alpha^{k-1}}{(k-1)!} = \alpha e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}$$
$$= \alpha e^{-\alpha}(e^{\alpha}) = \alpha$$
$$E[X^2] = \sum_{n=0}^{\infty} n^2 \frac{\alpha^n}{n!} e^{-\alpha} = e^{-\alpha}\alpha^2 \sum_{n=0}^{\infty} [n(n-1)+n] \frac{\alpha^{n-2}}{n!}$$
$$= e^{-\alpha}\alpha^2 \sum_{n=0}^{\infty} n(n-1)\frac{\alpha^{n-2}}{n!} + e^{-\alpha}\alpha \sum_{n=0}^{\infty} n \frac{\alpha^{n-1}}{n!}$$
$$= e^{-\alpha}\alpha^2 \frac{d^2}{d\alpha^2} e^{\alpha} + e^{-\alpha}\alpha \frac{d}{d\alpha} e^{\alpha} = [\alpha^2 + \alpha]$$
$$\text{Var}[X] = E[X^2] - E[X]^2 = \alpha^2 + \alpha - \alpha^2 = \alpha$$

Relation between Bernoulli, Binomial and Poisson RVs

A Poisson RV can be viewed as the limit of a sum of Bernoulli trials, i.e., as the limit of a binomial RV

– n Bernoulli trials, each with probability of success α/n

– K_n is the number of successes in n trials

– As $n \to \infty$, $P_{K_n}(k)$ converges to the PMF of a Poisson random variable with parameter α

– Poisson RV = sum of an infinite number of infinitesimal Bernoulli RVs

Proof

Since K_n is binomial, parameters $n, \alpha/n$, we have

$$P_{K_n}(k) = \binom{n}{k} (\alpha/n)^k (1-\frac{\alpha}{n})^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{\alpha^k}{k!}(1-\frac{\alpha}{n})^{n-k}$$

Note:

$$\lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1$$

$$(1-\frac{\alpha}{n})^{n-k} = \frac{(1-\frac{\alpha}{n})^n}{(1-\frac{\alpha}{n})^k}$$

$$\lim_{n\to\infty}(1-\frac{\alpha}{n})^n = e^{-\alpha} \qquad \lim_{n\to\infty}(1-\frac{\alpha}{n})^k = 1$$

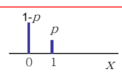
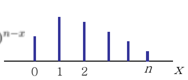
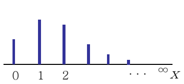
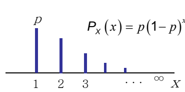
So,

$$\lim_{n\to\infty} P_{K_n}(k) = \frac{\alpha^k}{k!} e^{-\alpha}$$

Median and Mode of Special RVs

- Bernoulli RV, parameter 0.4
  – No median, mode = 0
- Geometric RV, parameter 0.5
  – Median is any number between 1 and 2, mode = 1
- Binomial distribution, parameters 2, 0.5:
  – Median = mode = 1

Summary of Special Discrete Random Variables

	PMF		Mean	Variance
Bernoulli	$P_X(1) = p$ $P_X(0) = 1 - p$ Number of successes in binary experiment p = probability of success		p	$p(1-p)$
Binomial	$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ Number of successes in n Bernoulli trials		np	$np(1-p)$
Poisson	$P_X(n) = \frac{\alpha^n}{n!} e^{-\alpha}, n \geq 0$ Number of successes in n Bernoulli trials as $p \to 0$, $n \to \infty$, and $np \to \alpha$		α	α
Geometric	$P_X(x) = p(1-p)^{x-1}$ Number Bernoulli trials needed to achieve one success	 $P_X(x) = p(1-p)^{x-1}$	$1/p$	$(1-p)/p^2$

Discrete RVs in MATLAB Statistics Toolbox

Distribution	PMF	CDF	Sample
Discrete uniform 1:N	unidpdf(X,N)	unidcdf(X,N)	unidrnd(N)
Bernoulli p	binopdf(X,1,p)	binocdf(X,1,p)	binornd(1,p)
Binomial n, p	binopdf(X,n,p)	binocdf(X,n,p)	binornd(n,p)
Geometric p	geopdf(X,p)	geocfd(X,p)	geornd(p)
Poisson α	poisspdf(X,α)	poisscdf(X,α)	poissrnd(α)

Example

CDF of a Binomial RV with parameters $n = 5$, $p = 0.53$

```
X = 5*linspace(-0.1,1.1);
Y = binocdf(X,5, 0.53);
stairs(X,Y)
```


MATLAB m-files on Web Site

MATLAB Functions

PMF	CDF	Random Sample
finitepmf(sx,p,x)	finitecdf(sx,p,x)	finiterv(sx,p,m)
bernoullipmf(p,x)	bernoullicdf(p,x)	bernoullirv(p,m)
binomialpmf(n,p,x)	binomialcdf(n,p,x)	binomialrv(n,p,m)
geometricpmf(p,x)	geometriccdf(p,x)	geometricrv(p,m)
pascalpmf(k,p,x)	pascalcdf(k,p,x)	pascalrv(k,p,m)
poissonpmf(alpha,x)	poissoncdf(alpha,x)	poissonrv(alpha,m)
duniformpmf(k,l,x)	duniformcdf(k,l,x)	duniformrv(k,l,m)

These and some others …

Table 2.1 (p. 88)
MATLAB Functions