ECE 340
Probabilistic Methods in Engineering
M/W 3-4:15

Lecture 8: Discrete RVs

Prof. Vince Calhoun

Review of Jean's Lectures

- Independence (2.5-2.6)
- Counting/Conditional Probability (3.1-3.2)
- Expected Value (3.3)
- Conditional PMF/Conditional EV (3.4)

Reading

- This class: Section 3.5
- Next class: Section 4.1-4.3

Outline

- Section 3.5
  - Discrete RV's
  - Examples
    - Bernoulli
    - Binomial
    - Geometric
    - Poisson

Discrete Random Variables

- Let X be a discrete RV that takes on the values: x_1, x_2, …, x_k, …
- We can define a function $g(x)$ of the discrete random variable X
- Although X is discrete, $g(x)$ can be continuous:

  $g(x)=ax+b$; $g(x)=x^2$; $g(x)=e^x$; $g(x)=\cos(x)$;…

Discrete Random Variables

- If X is a discrete RV that takes on the values: x_1, x_2, ..., x_k, ... then:

- $g(x) \Rightarrow g(x_1), g(x_2), \ldots, g(x_k), \ldots$

- Let $p_1=P[X=x_1]$, ... , $p_k=P[X=x_k]$,... then:

$$E[g(x)] = \sum_k g(x_k)\, p_k$$

Discrete Random Variable

- $g(x)=x \quad \Rightarrow E[g(x)]=m$ is the mean of X

$$E[g(x)] = E[X] = m = \sum_k x_k \, p_k$$

- $g(x)=(x-m)^2 \Rightarrow E[g(x)]=\sigma^2$ is the variance of X

$$E[g(x)] = E[(X-m)^2] = \sigma^2 = \sum_k (x_k - m)^2 \, p_k$$

$$\sigma^2 = \sum_k {x_k}^2 \, p_k - m^2 = E[X^2] - m^2$$

Examples of Discrete RVs

- **Bernoulli Random Variable**
  - Defined around a <u>single</u> event A
  - The outcome (of the underlying sample space S) is binary: the event A occurs <u>or</u> the event does not occur (i.e. A^c) $\Rightarrow$ Bernoulli trial
  - Transmit a packet: the packet is lost <u>or</u> it is not lost
  - Transmit a bit: the bit is received correctly <u>or</u> a bit-error event occurs

Examples of Discrete RVs

- Bernoulli Random Variable
  - The Bernoulli random variable I is defined by mapping the event A into a real-number and its complement A^c into another real-number
  - Traditionally: $I(A)=1$; $I(A^c)=0$
  - $P[I=1] = p$; $P[I=0] = 1-p$

Bernoulli Distribution

i. A Bernoulli experiment is a random experiment of which outcomes can be classified as one of 2 values. Success or failure; male or female; defective or non-defective.

ii. Bernoulli Distribution: $p(x) = f(x) = p^x(1-p)^{1-x}$, $x = 0, 1$

iii. This is a p.m.f, which describes a r.v. X which follows the Bernoulli distribution. This Bernoulli Distribution r.v. X has an associated mean & variance.
$\mu = E[X] = \Sigma xp(x) = \Sigma xp^x(1-p)^{1-x} = (0)(1-p) + (1)(p) = p$
$\sigma^2 = V(X) = \Sigma(x - p)^2p^x(1-p)^{1-x} = p^2(1-p) + (1-p)^2p = p(1 - p) = pq$

Examples of Discrete RVs

- **Binomial Random Variable**
  - Defined based on a series of n independent Bernoulli trials
  - The binomial RV X is the number of occurrences of the event *A* in the n trials

Be careful about the "definition of the event *A*"

- Transmit n packets: count the number of packets that are lost
- Transmit n bits: count the number of bits that are received correctly

Examples of Discrete RVs

- **Binomial Random Variable**

  - If the Bernoulli random variable is defined by the mapping: $I_j(A)=1$ & $I_j(A^c)=0$ for $j=1, 2, \ldots, k, \ldots, n$ then the binomial RV X can be expressed:

$$X = \sum_{j=1}^{n} I_j$$

  - Therefore, $S_X=\{0, 1, 2, 3, \ldots, k, \ldots, n\}$

Examples of Discrete RVs

- Binomial Random Variable
  - Let $P[I_j(A)=1] = p$, $P[I_j(A^c)=0] = 1-p$ $\forall$ $j=1, 2, \ldots, n$
  - Then for the binomial RV: $X = \sum_{j=1}^{n} I_j$

  the probability that $X=k$, for $k=0, 1, \ldots, n$:

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

Binomial Distribution

i. Note: We are only interested in the total # of successes and NOT in the order of those successes. Order is irrelevant.

ii. If we let the r.v. X equals to the number of observed successes in n Bernoulli Trials, the possible values of X are 0, 1, 2, ..., n.

If x successes occur (x = 0, 1, ..., n) then n – x failures occur. The number of ways of selecting x positions for the x successes in the n trials is: $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

with this fact, the p.m.f of a binomial r.v.X is:

$$b(x;n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, x = 0,1,\ldots,n \\ 0, elsewhere \end{cases}$$

Binomial Distribution

i. Def: for X~bin(n, p) the cumulative distribution function will be denoted by:

P(X ≤ x) = B(x; n, p) = $\sum_{y=0}^{x} b(y; n, p)$

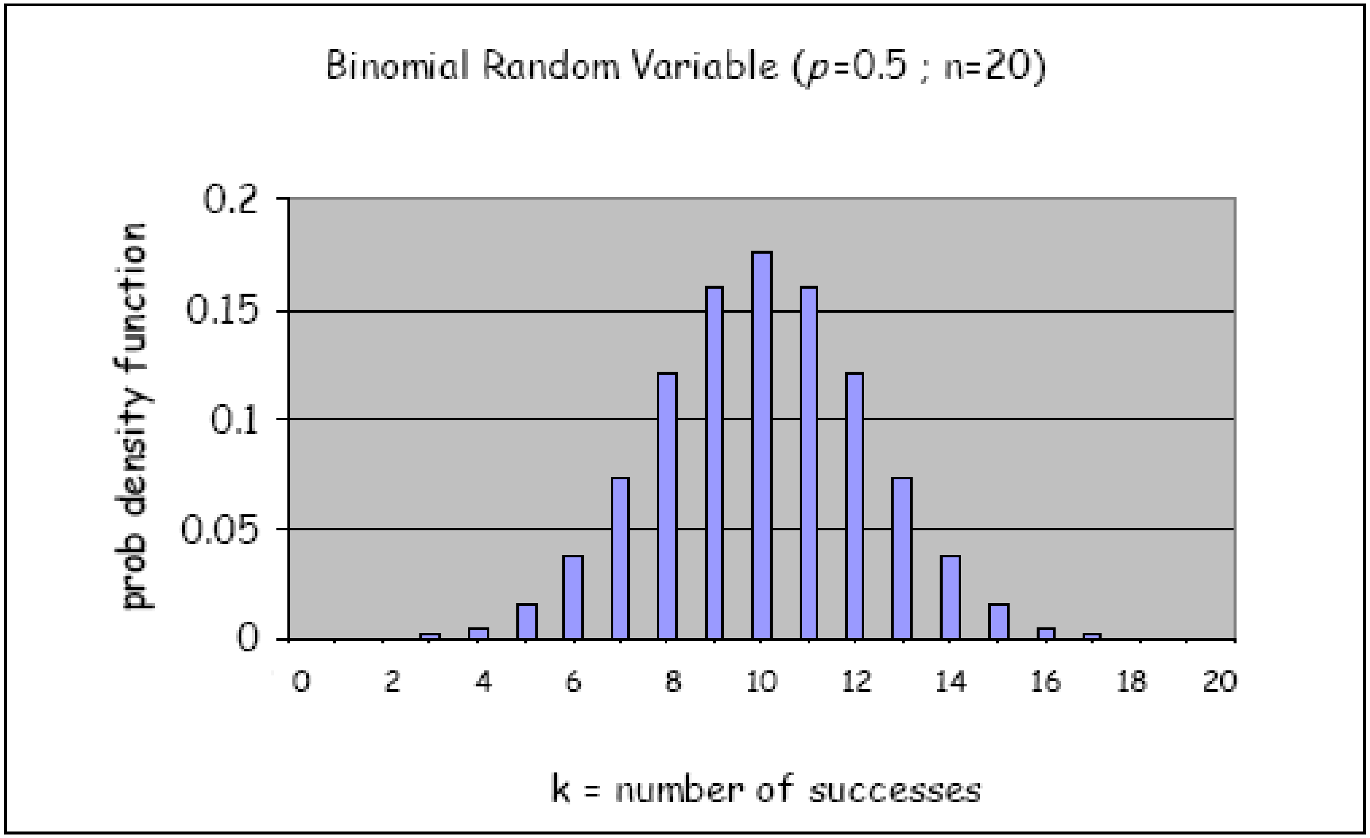
ii. For n = 1, the binomial distribution becomes the Bernoulli distribution.

iii. If X ~ bin(n, p)

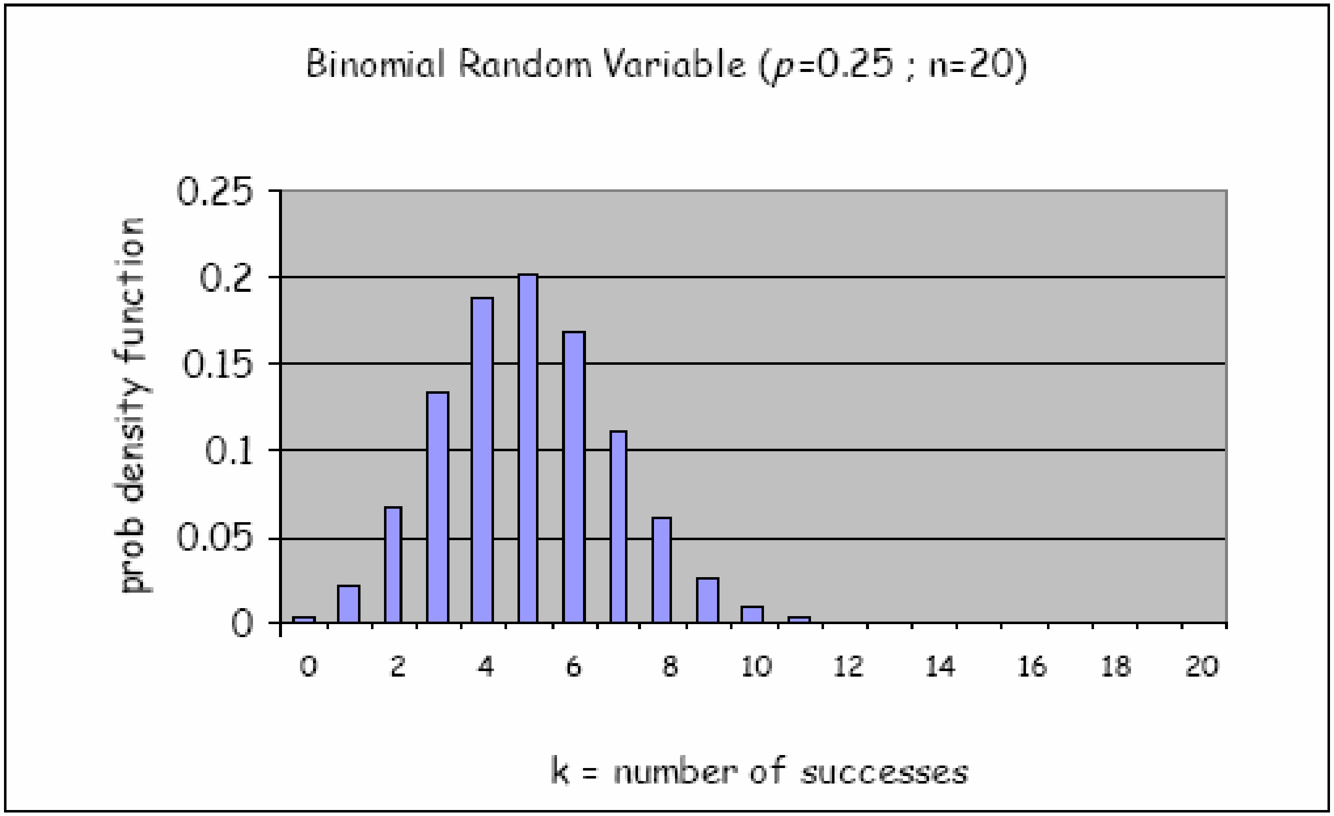
$$E[X] = \sum_{x=0}^{n} xp(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = np$$

$$V(X) == E[X^2] - (E[X])^2 = \sum_{x=0}^{n} x^2 p(x) - (np)^2 == npq$$

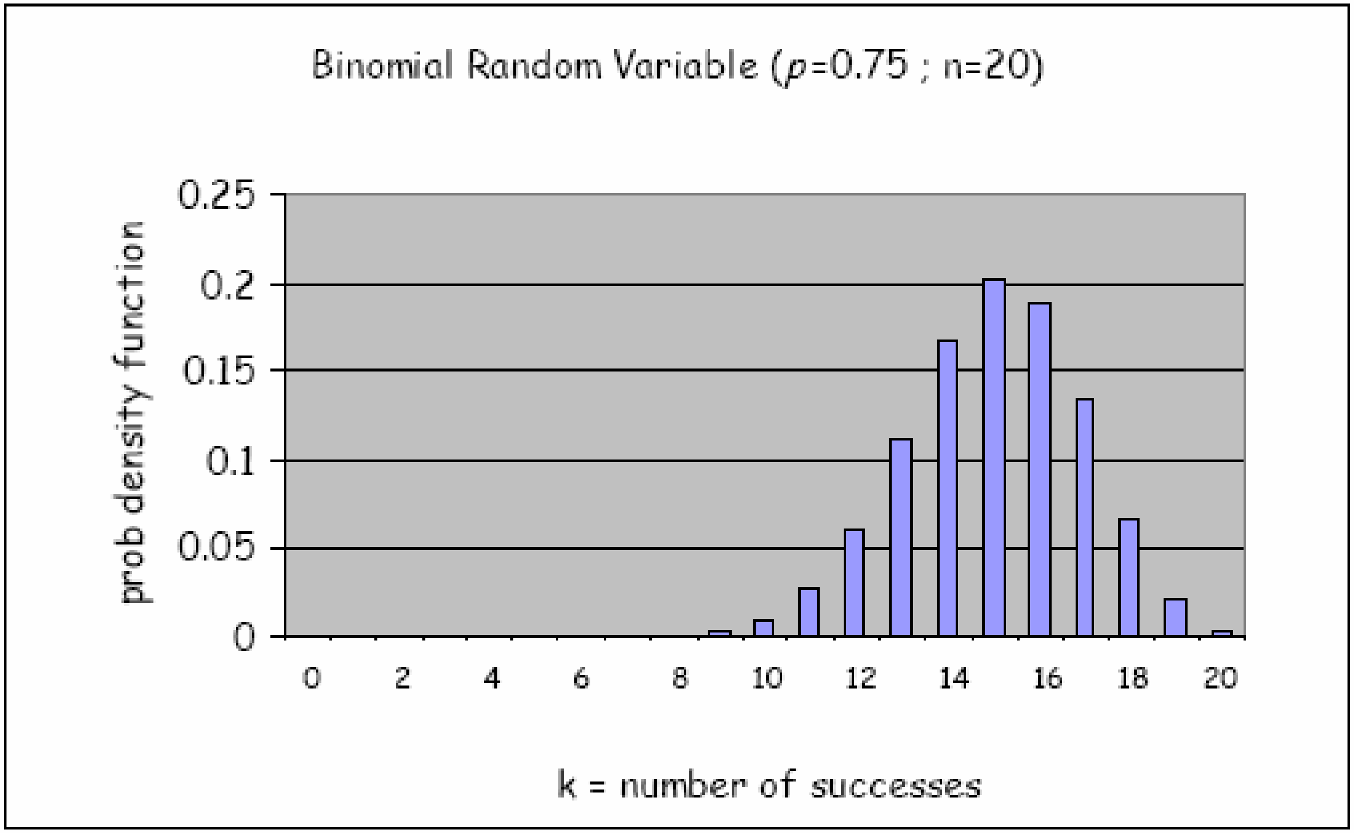
Binomial Random Variable



Binomial Random Variable



Binomial Random Variable



Binomial Distribution

- **A process uses an inspection plan that calls for a sample of 5 units to be checked before shipment. One failure rejects the lot. If there are 10% defective, what is the probability of lot rejection?**
- **In this example, *p* = 0.10, and *n* = 5, giving the following *pmf*:**

$$P(X=x)=\frac{5!}{x!(5-x)!}(0.10)^{x}(0.90)^{(5-x)}$$

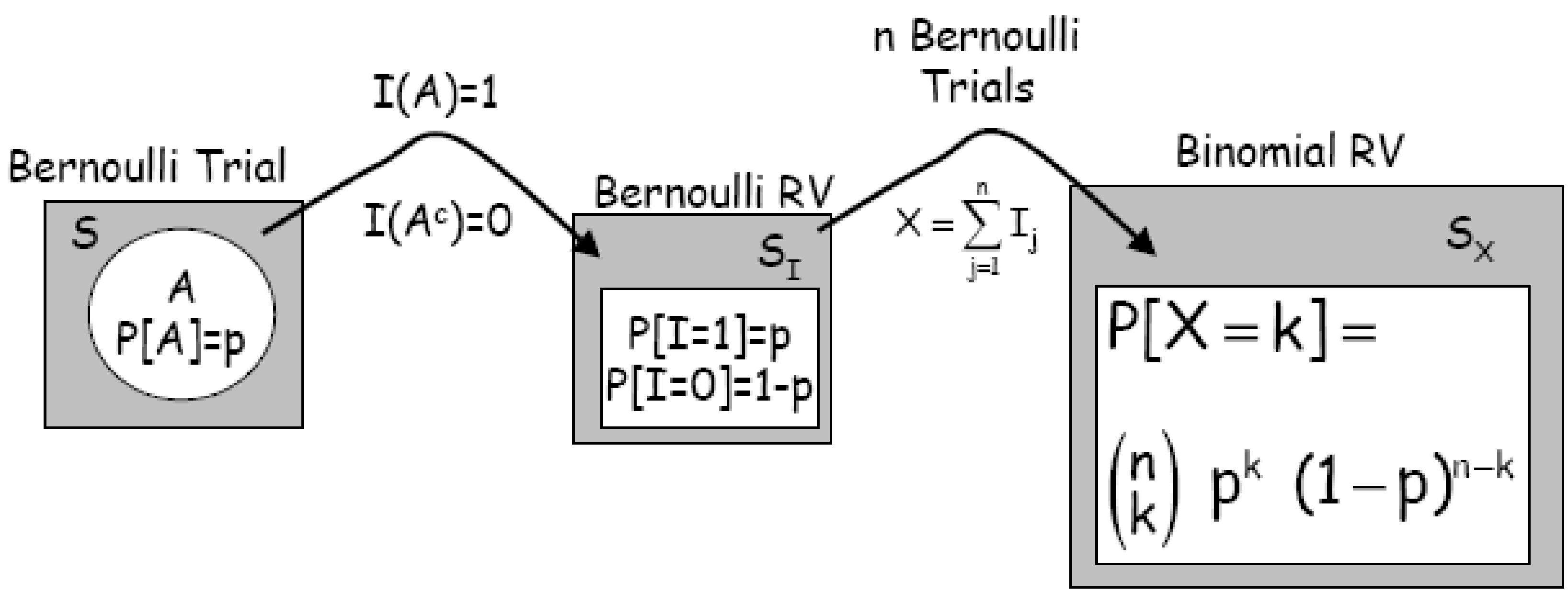
Binomial Distribution

- This yields the following distribution table for the possible results:

Failures	Pmf value	CDF value
0	0.59049	0.59049
1	0.32805	0.91854
2	0.07290	0.99144
3	0.00810	0.99954
4	0.00045	0.99999
5	0.00001	1.00000

- There is a 59% chance that failing lots would be sent out.

Bernoulli Trials and RVs



Geometric Probability Distribution

The geometric distribution is different from the binomial distribution in that the r.v. X is the number of trials on which the first success occurs (rather than the number of the successes that occur in n trials).

The experiment can end with the 1^{st} trial (if successful) or it can go on indefinitely

Discrete RVs: *Geometric*

- Geometric random variable M
- M is the number of Bernoulli trials <u>until</u> the first occurrence of the event A
- The number of: trials resulting in the event A^c ("failures") <u>plus</u> the first occurrence of A ("success")
- E.g., Sending a packet over the Internet: Keep re-transmitting until the packet makes it!

Discrete RVs: Geometric

- Geometric random variable M has an infinitely countable possible values:

$$S_M=\{1,\ 2,\ 3,\ \dots\ k,\ \dots\ \}$$

- The probability that M takes on a value k:

$$P[M=k] = (1-p)^{k-1}\, p$$

The "success" trial is included in counting M

- Be careful, sometime M is defined such that:

$$P[M=k] = (1-p)^{k}\, p$$

The "success" trial is not included in counting M

Geometric Probability Distribution

For the geometric distribution

$$E[X] = \frac{1}{p}$$

$$V(X) = \sigma^2 = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Note: p represents a number between 0 and 1 and is a parameter of the probability distribution family.

Discrete RVs: Memory-less RV

- Given that after j trials the event *A* did not occur (i.e. j "failures"), what is the probability that we need at least other k trials <u>until</u> *A* occurs?

- If we have a "memory-less" experiment, the probability that "we need at least other k trials <u>until</u> we have a 'successful' trial" is independent of the past:

$$P[\ \{ M \geq k+j \} / \{M > j\}\ ] = P[\ M \geq k\ ]$$

Discrete RVs: Memory-less RV

- The probability that a geometric RV $M \geq k$:

$$P[M \geq K] = \sum_{i=k}^{\infty} (1-p)^{i-1}\, p = (1-p)^{k-1}$$

- Show that: a RV M satisfies the memory-less property if-and-only-if M is a geometric RV

Poisson Probability Distribution

Consider a discrete r.v. which is often useful when dealing with the number of occurrences of an event over a specified interval of time. Suppose we want to find the probability distribution of the accidents at the intersection of First and Santa Clara streets during a one-week period. Since the accident event is discrete and countable (from zero to whatever), we use the Poisson distribution to model the accident probability.

The R.V. we are interested in is the number of accidents.

Discrete RVs: Poisson

- Poisson random variable N is the number of occurrences of a "completely random event" over a unit of time
- $S_N = \{0, 1, 2, 3, \ldots\}$
- E.g., the number of packets that arrive at a network router over one second; number of customers arriving at a queue between noon and 1pm; etc.

Discrete RVs: Poisson

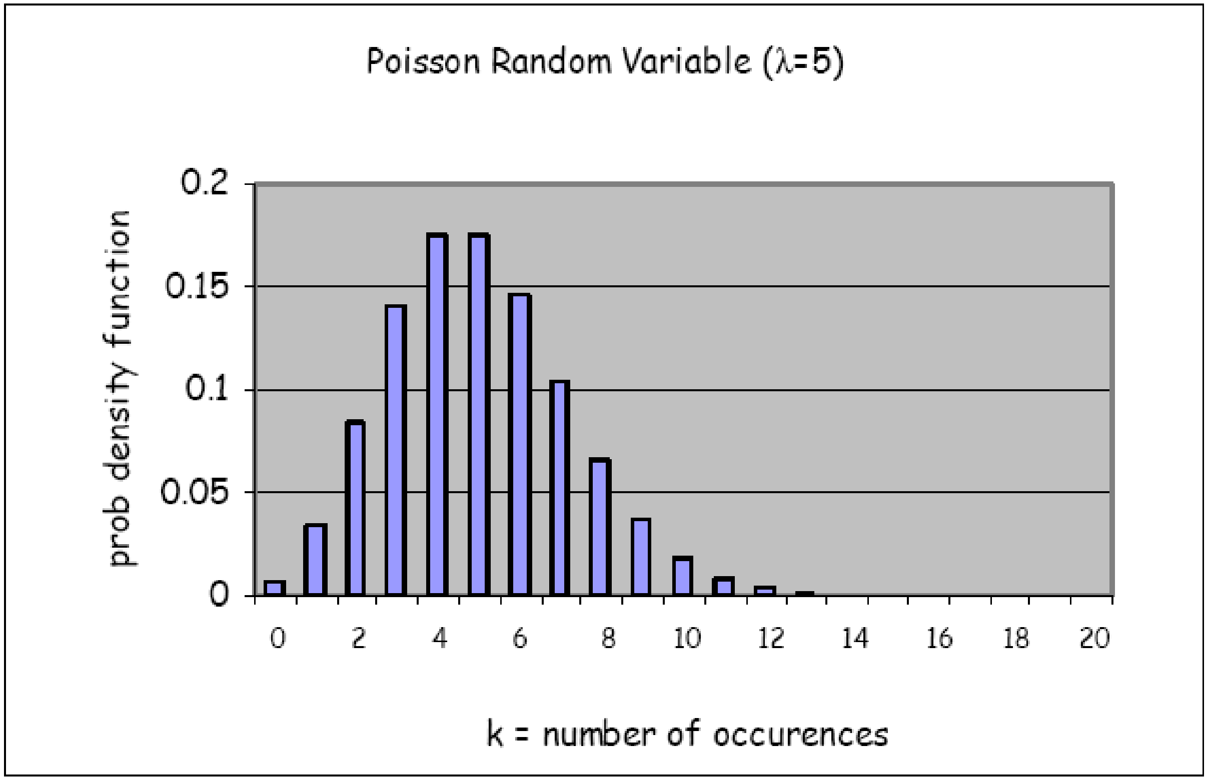
- A Poisson RV is characterized by a parameter λ:

  λ=the average number of occurrences per a unit of time

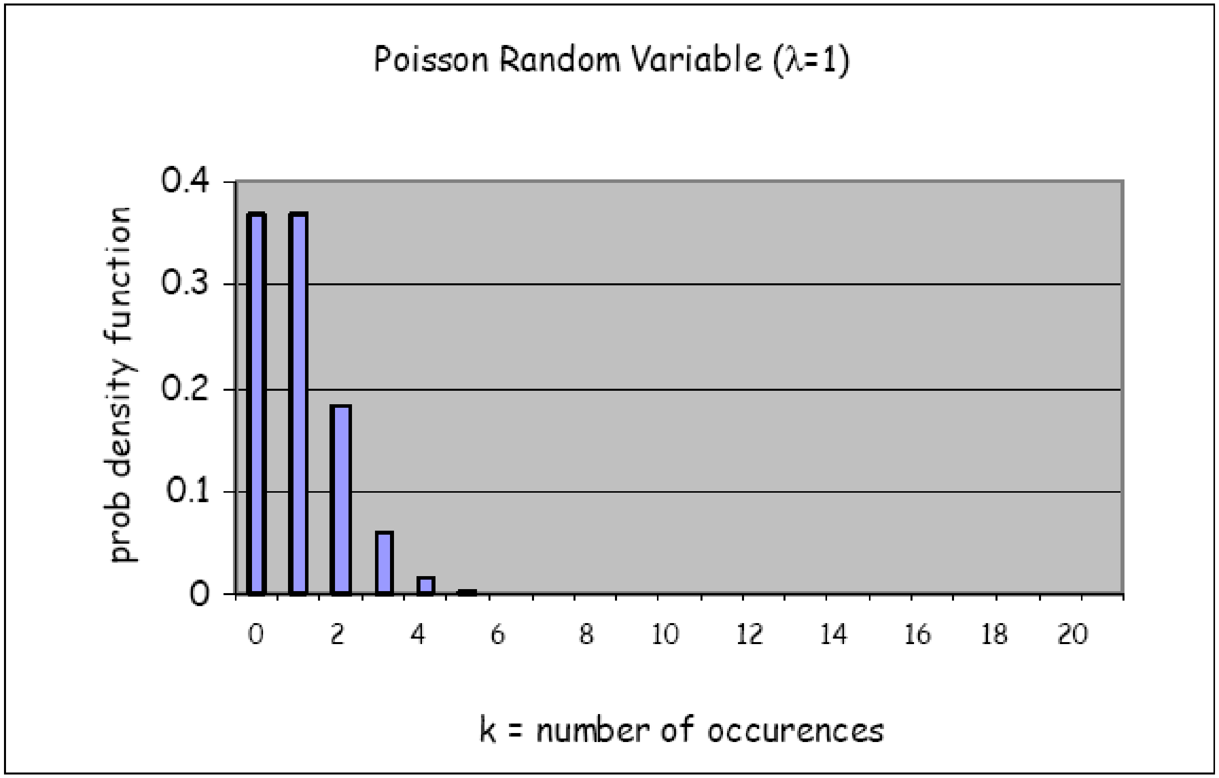
- The probability that a Poisson RV N takes on a value k (over a unit of time):

$$P[N = k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

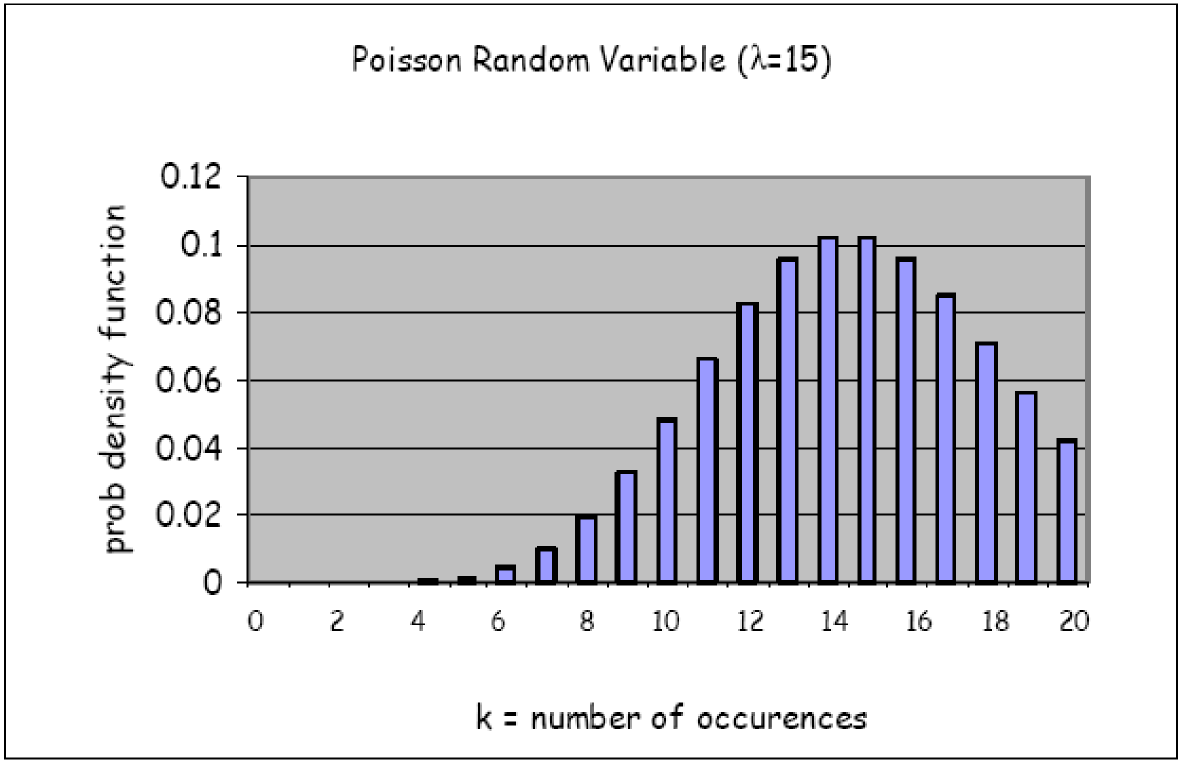
Discrete RVs: Poisson



Discrete RVs: Poisson



Discrete RVs: Poisson



Discrete RVs: Poisson Moments

- The expected value (mean) of a Poisson RV:

$$E[N] = \sum_{k=0}^{\infty} k\, p_k = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = ?$$

- The second moment of a Poisson RV:

$$E[N^2] = \sum_{k=0}^{\infty} k^2\, p_k = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = ?$$

Poisson Probability Distribution

i. The Poisson distribution is built on the foundations of the binomial distribution, thus $\lambda = \mu = np$

ii. If X has a Poisson distribution with parameter λ, then
$E[X] = \lambda$
$V(X) = \lambda$
mean and variance are the same

Discrete RVs: Poisson Moments

- To derive the moments of the Poisson random variable, you need to remember some old results from calculus.

  The infinite series expansion of e^{λ}:

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

Poisson Probability Distribution

- A manufacturer checks for contamination on their storage disks. The mean value is 0.1 contaminants per square centimeter, with a disk surface of 100 square centimeters. What is the probability of five or more contaminants on the disks?
- The expected value per disk is:
- 100 * 0.10 = 10 contaminants per disk

Poisson Probability Distribution

- The question asked for 5 or more, which can be calculated by difference, e.g.
- 1 – P(0) – P(1) – P(2) – P(3) – P (4)
- lambda = 10, so the basic relation is:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-10}10^{x}}{x!}$$

Poisson Probability Distribution

- Tabulating the results and subtracting from 1 gives:

X=	P(X=x)
0	0.0000454
1	0.0004540
2	0.0022700
3	0.0075667
4	0.0189161

- This totals 0.0292528, leaving the probability as 0.9707472 that 5 or more contaminants will be found

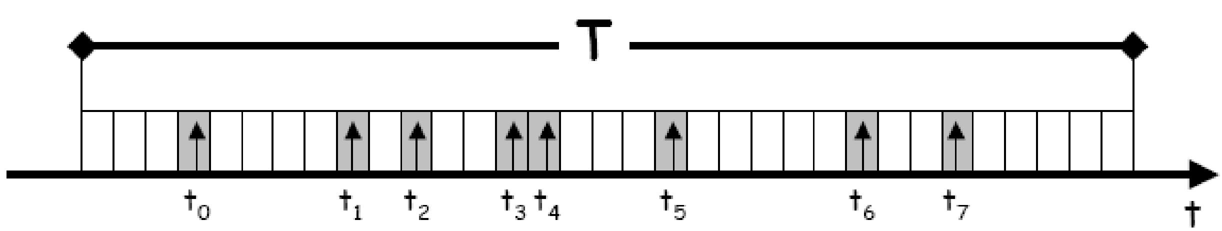
Discrete RVs: Poisson

- λ=the <u>average</u> number of occurrences per <u>one unit of time</u>
- The probability that a Poisson RV N takes on a value k over t units of time:

$$P[N = k \text{ in } t \text{ time units}] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Discrete RVs: Poisson

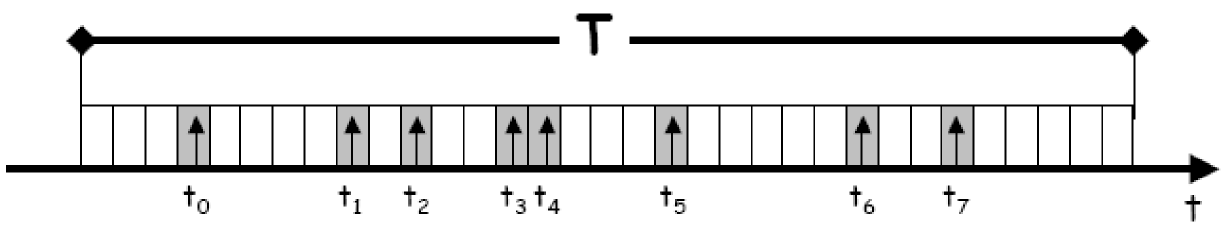
- If we divide a "Poisson time unit" T into a very-large number (e.g. n) of very-small time intervals ("ticks") such that:
  - The event A can occur "at most" once in any "tick"
- This can be viewed as n independent Bernoulli trials with $P[A]=p=\lambda/n$



Discrete RVs: Poisson

- The probability that the event A occurs k times in these n trials can be computed using the binomial pmf (with $p=\lambda/n$):

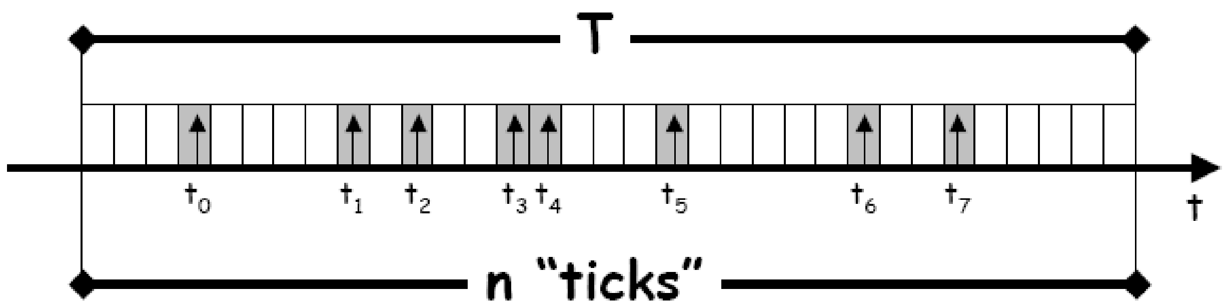
$$P[X=k]=\binom{n}{k} p^k (1-p)^{n-k}$$



Discrete RVs: Poisson

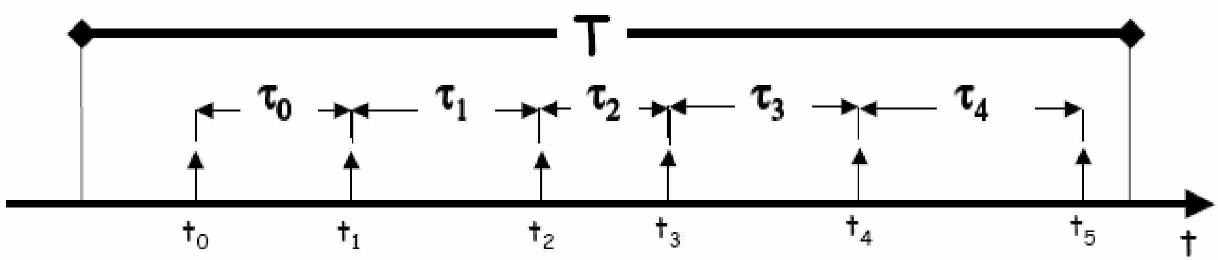
- This leads to the following approximation of the binomial pmf (with $p=\lambda/n$) by the Poisson pmf:

$$\lim_{n\to\infty}\binom{n}{k}(\lambda/n)^k\,(1-(\lambda/n))^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$$



Examples of Random Variables

- Let us define the variable $\tau_n = t_{n+1} - t_n$
- Let $\tau = \{\tau_n\ ,\ n = 0,\ 1,\ 2,\ \ldots.\}$
- Therefore, τ measures the distance (in time) between the occurrences of a Poisson random event
- Is τ a random variable?



Common Families of Discrete Probability Distribution

Bernoulli B(1,p)	$p(x) = f(x) = p^x(1-p)^{1-x}$, $x = 0, 1$ $M(t) = 1 - p + pe^t$ $\mu = p$, $\sigma^2 = p(1-p) = pq$
Binomial B(n, p)	$p(x) = f(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, x = 0,1\ldots,n$ $M(t) = (1 - p + pe^t)^n$ $\mu = p$, $\sigma^2 = p(1-p) = pq$
Geometric	$p(x) = f(x) = (1-p)^{x-1}p$, $x = 1, 2, \ldots$ $M(t) = \frac{Pe^t}{1-(1-p)e^t}, t < -\ln(1-p)$ $\mu = \frac{1}{p}$, $\sigma^2 = \frac{1-p}{p^2} = \frac{q}{p^2}$
Poisson	$p(x) = f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, $x = 0,1, \ldots$ $M(t) = e^{\lambda(e^t-1)}$ $\mu = \lambda$, $\sigma^2 = \lambda$

Generation of discrete RV’s in Matlab

- rand
- poissrnd
- binornd