UCF Stands For Opportunity

CDA6530: Performance Models of Computers and Networks

Chapter 2: Review of Practical Random Variables

SCHOOL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

Two Classes of R.V.

- Discrete R.V.
  - Bernoulli
  - Binomial
  - Geometric
  - Poisson
- Continuous R.V.
  - Uniform
  - Exponential, Erlang
  - Normal
- Closely related
  - Exponential $\leftarrow\rightarrow$ Geometric
  - Normal $\leftarrow\rightarrow$ Binomial, Poisson

Definition

- Random variable (R.V.) X:
  - A function on sample space
  - X: S $\rightarrow$ R
- Cumulative distribution function (CDF):
  - Probability distribution function (PDF)
  - Distribution function
  - $F_X(x) = P(X \leq x)$
  - Can be used for both continuous and discrete random variables

- Probability density function (pdf):
  - Used for continuous R.V.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \qquad f_X(x) = \frac{dF_X(x)}{dx}$$

- Probability mass function (pmf):
  - Used for discrete R.V.
  - Probability of the variable exactly equals to a value

$$f_X(x) = P(X = x)$$

Bernoulli

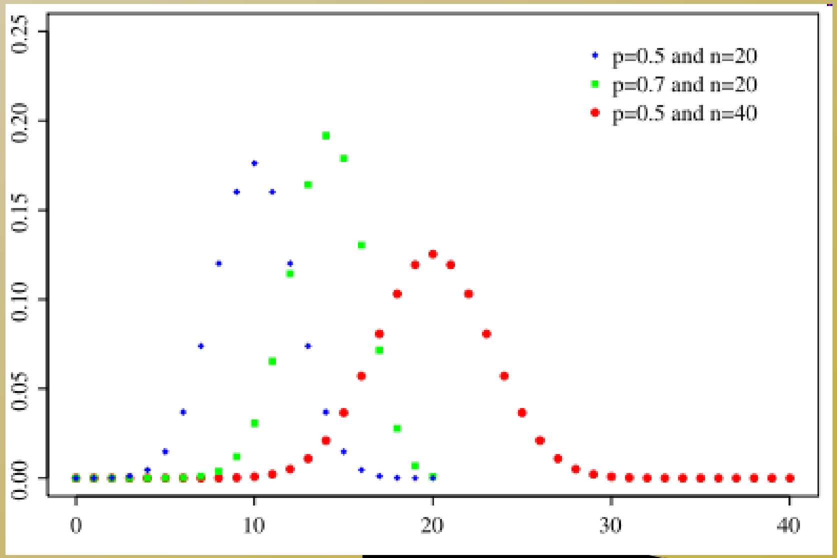
- A trial/experiment, outcome is either "success" or "failure".
  - X=1 if success, X=0 if failure
  - P(X=1)=p, P(X=0)=1-p
- Bernoulli Trials
  - A series of independent repetition of Bernoulli trial.

Binomial

- A Bernoulli trials with n repetitions
- Binomial: X = No. of successes in n trails
  - X$\sim$ B(n, p)

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

where $$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$



Binomial Example (1)

- A communication channel with (1-p) being the probability of successful transmission of a bit. Assume we design a code that can tolerate up to *e* bit errors with *n* bit word code.
- Q: Probability of successful word transmission?
- Model: sequence of bits trans. follows a Bernoulli Trails
  - Assumption: each bit error or not is independent
  - P(Q) = P(e or fewer errors in n trails)

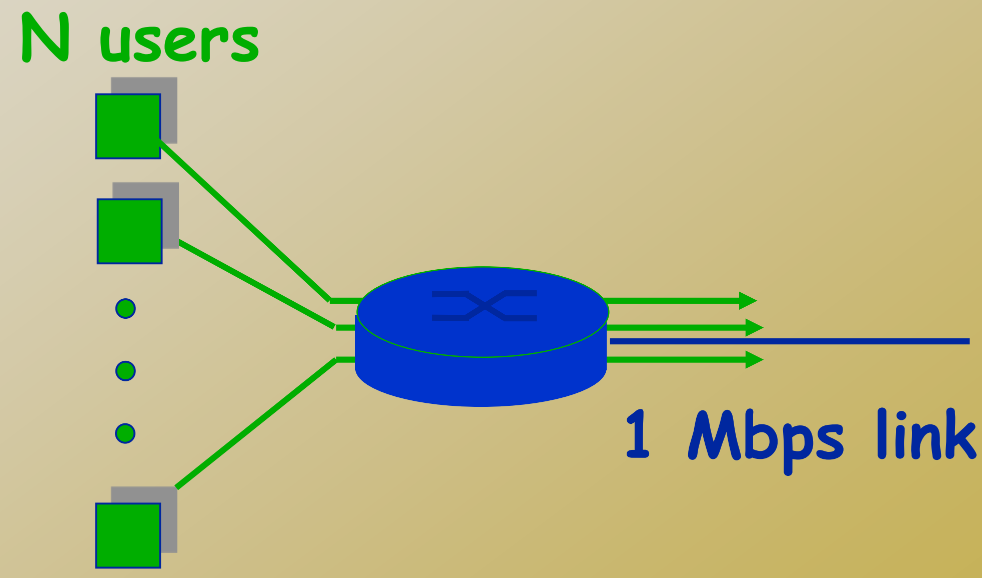
$$= \sum_{i=0}^{e} f(i; n, p)$$

$$= \sum_{i=0}^{e} \binom{n}{i} p^i (1-p)^{n-i}$$

Binomial Example (2)
---- Packet switching versus circuit switching

Packet switching allows more users to use network!

- 1 Mb/s link
- each user:
  - 100 kb/s when "active"
  - active 10% of time
- circuit-switching:
  - 10 users
- packet switching:
  - with 35 users, prob. of > 10 active less than .0004

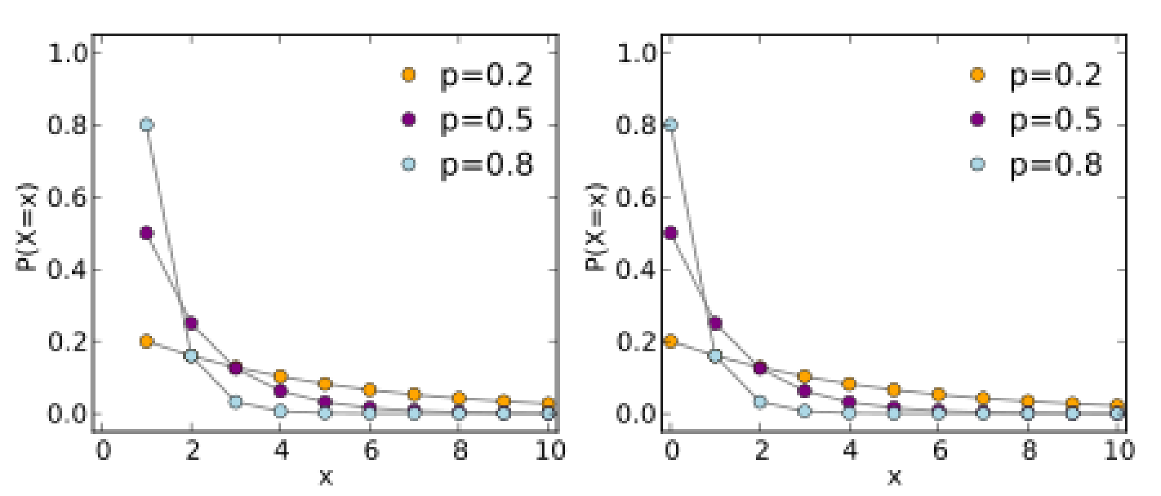


Q: how did we know 0.0004?

Geometric

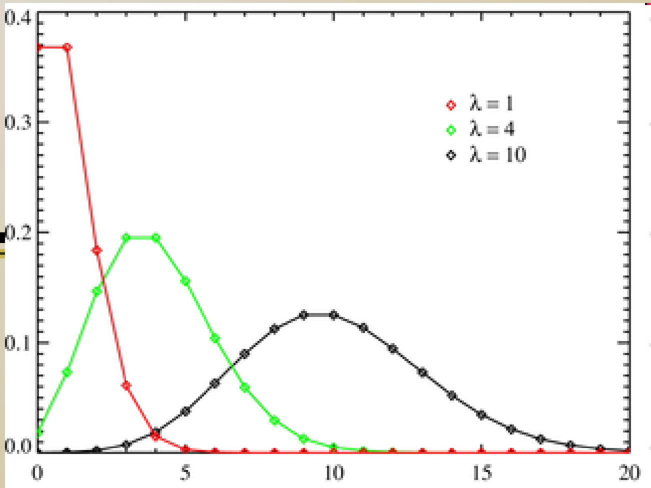
- Still about Bernoulli Trails, but from a different angle.
- X: No. of trials until the first success
- Y: No. of failures until the first success
- $P(X=k) = (1-p)^{k-1}p$ $\qquad$ $P(Y=k)=(1-p)^{k}p$

X



Y

Poisson



- Limiting case for Binomial when:
  - n is large and p is small
    - n>20 and p<0.05 would be good approximation
      - λ=np is fixed, success rate
- X: No. of successes in a time interval (n time units)

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

- Remember that X follows Binomial distr., so real value of P(X=k) is:

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

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

- Why the approximation is accurate?

Poisson

- Many natural systems have this distribution
  - The number of phone calls at a call center per minute.
    - Tens of thousands of customers out there.
    - Each customer has very tiny probability to call at a specific minute period.
  - The number of times a web server is accessed per minute.
  - The number of mutations in a given stretch of DNA after a certain amount of radiation.

Continous R.V - Uniform

- X: is a uniform r.v. on (α, β) if

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{if} \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

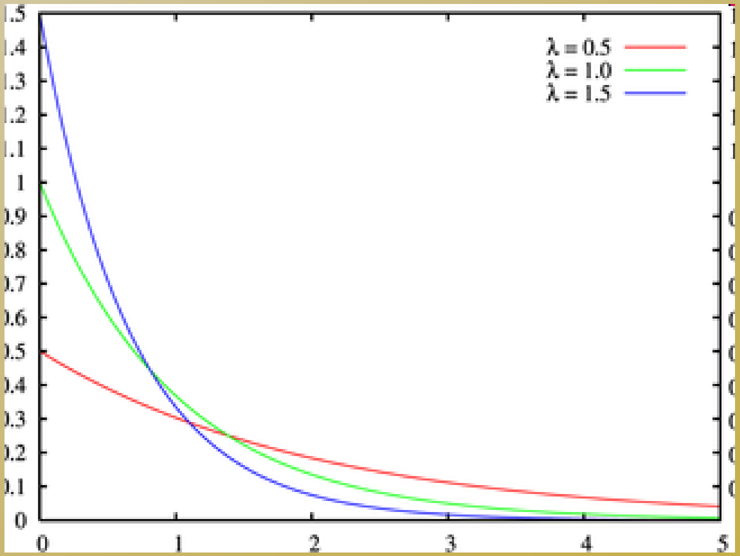
- Uniform r.v. is the basis for simulation other distributions
  - Introduce later

Exponential

- r.v. X:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

- $F_X(x) = 1 - e^{-\lambda x}$
- Very important distribution
  - Memoryless property (explained a bit later)
  - Corresponding to geometric distr.

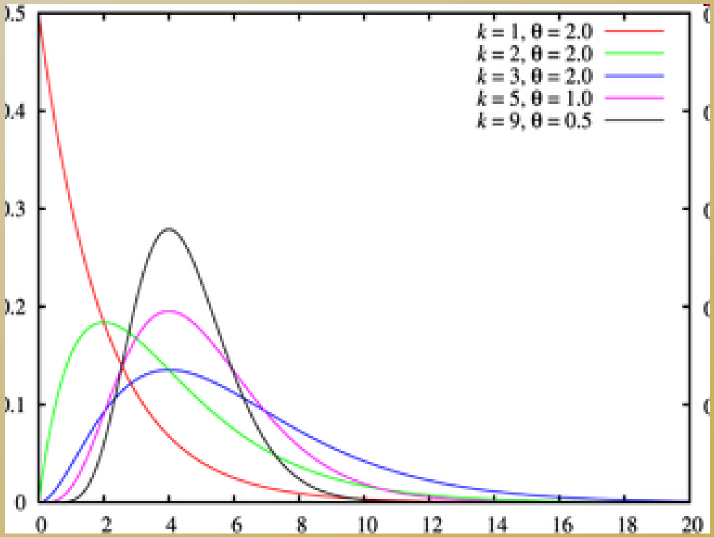


Erlang

- r.v. X (k-th Erlang):

$$f(x; k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \quad \text{for } x, \lambda \geq 0.$$

- K-th Erlang is the sum of k Exponential distr.

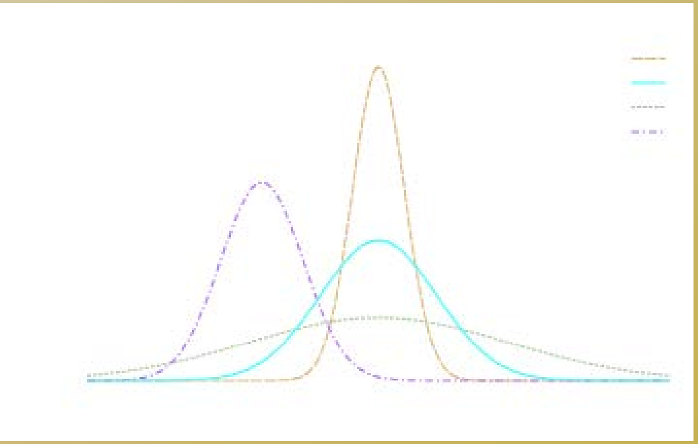


Normal

- r.v. X:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < \infty$$

- Corresponding to Binomial and Poisson distributions

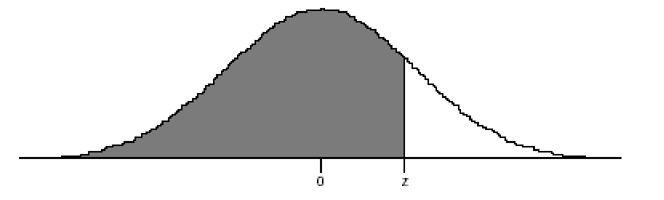


Normal

- If $X \sim N(\mu, \sigma^2)$, then
  - r.v. $Z=(X-\mu)/\sigma$ follows standard normal N(0,1)
  - $P(Z<x)$ is denoted as $\Phi(x)$
    - $\Phi(x)$ value can be obtained from standard normal distribution table (next slide)
  - Used to calculate the distribution value of a normal random variable $X \sim N(\mu, \sigma^2)$
    - $P(X<\alpha) = P(Z < (\alpha-\mu)/\sigma)$

      $= \Phi((\alpha-\mu)/\sigma)$

Standard Normal Distr. Table



z	F(x)	z	F(x)	z	F(x)
-2.5	0.006	-1	0.159	0.5	0.691
-2.4	0.008	-0.9	0.184	0.6	0.726
-2.3	0.011	-0.8	0.212	0.7	0.758
-2.2	0.014	-0.7	0.242	0.8	0.788
-2.1	0.018	-0.6	0.274	0.9	0.816
-2	0.023	-0.5	0.309	1	0.841
-1.9	0.029	-0.4	0.345	1.1	0.864
-1.8	0.036	-0.3	0.382	1.2	0.885
-1.7	0.045	-0.2	0.421	1.3	0.903
-1.6	0.055	-0.1	0.46	1.4	0.919
-1.5	0.067	0	0.5	1.5	0.933
-1.4	0.081	0.1	0.54	1.6	0.945
-1.3	0.097	0.2	0.579	1.7	0.955
-1.2	0.115	0.3	0.618	1.8	0.964
-1.1	0.136	0.4	0.655	1.9	0.971

- $P(X<x) = \Phi(x)$
  - $\Phi(-x) = 1- \Phi(x)$ why?
  - About 68% of the area under the curve falls within 1 standard deviation of the mean.
  - About 95% of the area under the curve falls within 2 standard deviations of the mean.
  - About 99.7% of the area under the curve falls within 3 standard deviations of the mean.

Normal Distr. Example

- An average light bulb manufactured by Acme Corporation lasts 300 days, 68% of light bulbs lasts within 300+/- 50 days. Assuming that bulb life is normally distributed.
  - Q1: What is the probability that an Acme light bulb will last at most 365 days?
  - Q2: If we installed 100 new bulbs on a street exactly one year ago, how many bulbs still work now on average? What is the distribution of the number of remaining bulbs?
- Step 1: Modeling
  - $X \sim N(300, 50^2)$ $\mu=300$, $\sigma=50$. Q1 is $P(X \leq 365)$
    define $Z = (X-300)/50$, then Z is standard normal
  - For Q2, # of remaining bulbs, Y, is a Bernoulli trial with 100 repetitions with small prob. $p = [1- P(X \leq 365)]$
    - Y follows Poisson distribution (approximated from Binomial distr.)
    - $E[Y] = \lambda = np = 100 * [1- P(X \leq 365)]$

Memoryless Property

- Memoryless for Geometric and Exponential
- Easy to understand for Geometric
  - Each trial is independent → how many trials before hitting a target does not depend on how many times I have missed before.
  - X: Geometric r.v., $P_X(k)=(1-p)^{k-1}p$;
  - Y: Y=X-n No. of trials given we failed first n times
  - $P_Y(k) = P(Y=k|X>n)=P(X=k+n|X>n)$

$$= \frac{P(X=k+n, X>n)}{P(X>n)} = \frac{P(X=k+n)}{P(X>n)}$$

$$= \frac{(1-p)^{k+n-1}p}{(1-p)^n} = p(1-p)^{k-1} = P_X(k)$$

- pdf: probability density function
  - Continuous r.v. $f_X(x)$
- pmf: probability mass function
  - Discrete r.v. X: $P_X(x)=P(X=x)$
  - Also denoted as $P_X(x)$ or simply $P(x)$

Mean (Expectation)

- Discrete r.v. X
  - $E[X] = \sum kP_X(k)$
- Continous r.v. X
  - $E[X] = \int_{-\infty}^{\infty} kf(k)dk$
- Bernoulli: $E[X] = 0(1-p) + 1\cdot p = p$
- Binomial: $E[X]=np$ (intuitive meaning?)
- Geometric: $E[X]=1/p$ (intuitive meaning?)
- Poisson: $E[X]=\lambda$ (remember $\lambda=np$)

Mean

- Continuous r.v.
  - Uniform: E[X]= $(\alpha+\beta)/2$
  - Exponential: E[X]= $1/\lambda$
    - K-th Erlang E[X] = k/λ
  - Normal: E[X]=μ

Function of Random Variables

- R.V. X,   R.V. Y=g(X)
- Discrete r.v. X:
  - E[g(X)] = $\sum$ g(x)p(x)
- Continuous r.v. X:
  - E[g(X)] = $$\int_{-\infty}^{\infty} g(x)f(x)dx$$

- Variance: Var(X) = E[(X-E[X])^2]
  = E[X^2] – (E[X])^2

UCF Stands For Opportunity

SCHOOL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

Joint Distributed Random Variables

- $F_{XY}(x,y)=P(X\leq x, Y\leq y)$
- $F_{XY}(x,y)=F_X(x)F_Y(y)$ if X and Y are independent
- $F_{X|Y}(x|y) = F_{XY}(x,y)/F_Y(y)$

- $E[\alpha X +\beta Y]=\alpha E[X]+\beta E[Y]$
- If X, Y independent
  - $E[g(X)h(Y)]=E[g(X)]\cdot E[h(Y)]$
- Covariance
  - Measure of how much two variables change together
  - $Cov(X,Y)=E[(X-E[X])(Y-E[Y])]$
    $= E[XY] – E[X]E[Y]$
  - If X and Y independent, $Cov(X,Y)=0$

Limit Theorems - Inequality

- Markov's Inequality
  - r.v. $X \geq 0$: $\forall\ \alpha>0$, $P(X \geq \alpha) \leq E[X]/\alpha$
- Chebyshev's Inequality
  - r.v. X, $E[X]=\mu$, $Var(X)=\sigma^2$
  - $\forall\ k>0$, $P(|X-\mu| \geq k) \leq \sigma^2/k^2$
- Provide bounds when only mean and variance known
  - The bounds may be more conservative than derived bounds if we know the distribution

Inequality Examples

- If α=2E[X], then P(X$\geq\alpha$)$\leq$ 0.5
- A pool of articles from a publisher. Assume we know that the articles are on average 1000 characters long with a standard deviation of 200 characters.
- Q: what is the prob. a given article is between 600 and 1400 characters?
- Model: r.v. X: μ=1000, σ=200, k=400 in Chebyshev's
- P(Q) = 1- P(|X-μ|$\geq$ k)
  $\geq$ 1- $(\sigma/k)^2$ =0.75

- If we know X follows normal distr.:
  - The bound will be tigher
  - 75% chance of an article being between 760 and 1240 characters long

Strong Law of Large Number

- i.i.d. (independent and identically-distributed)
- X_i: i.i.d. random variables, $E[X_i]=\mu$

With probability 1,
$(X_1+X_2+\cdots +X_n)/n \rightarrow \mu$, as $n \rightarrow \infty$

Foundation for using large number of simulations to obtain average results

Central Limit Theorem

- X_i: i.i.d. random variables, $E[X_i]=\mu$ $Var(X_i)=\sigma^2$
- Y= $\dfrac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$
- Then, $Y \sim N(0,1)$ as $n \rightarrow \infty$
  - The reason for why normal distribution is everywhere
  - Sample mean $\bar{X}$ is also normal distributed
- Sample mean

$$\bar{X} = \sum_{i=1}^{n} X_i/n$$

$$E[\bar{X}] = \mu$$

$$Var(\bar{X}) = \sigma^2/n$$

What does this mean?

Example

- Let X_i, $i=1,2,\cdots, 10$ be i.i.d., X_i is uniform distr. (0,1). Calculate $P(\sum_{i=1}^{10} X_i > 7)$

- $E[X_i]=0.5$, $Var(X_i)=1/12$

$$P(\sum_{i=1}^{10} X_i > 7) = P(\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{10(1/12)}} > \frac{7-5}{\sqrt{10(1/12)}})$$

$$\approx 1 - \Phi(2.2) = 0.0139$$

$\Phi(x)$: prob. standard normal distr. $P(X < x)$

Conditional Probability

- Suppose r.v. X and Y have joint pmf p(x,y)
  - $p(1,1)=0.5$, $p(1,2)=0.1$, $p(2,1)=0.1$, $p(2,2)=0.3$
  - Q: Calculate the pmf of X given that Y=1

- $p_Y(1)=p(1,1)+p(2,1)=0.6$
- X sample space {1,2}
- $p_{X|Y}(1|1) = P(X=1|Y=1) = P(X=1, Y=1)/P(Y=1)$

  $= p(1,1)/p_Y(1) = 5/6$

- Similarly, $p_{X|Y}(2,1) = 1/6$

UCF Stands For Opportunity

SCHOOL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

Expectation by Conditioning

- r.v. X and Y. then E[X|Y] is also a r.v.
- Formula: E[X]=E[E[X|Y]]
  - Make it clearer, $E_X[X]= E_Y[\, E_X[X|Y] \,]$
  - It corresponds to the "law of total probability"
    - $E_X[X]= \sum E_X[X|Y=y] \cdot P(Y=y)$
    - Used in the same situation where you use the law of total probability

Example

- r.v. X and N, independent
- $Y=X_1+X_2+\cdots+X_N$

- Q: compute E[Y]?

Example 1

- A company's network has a design problem on its routing algorithm for its core router. For a given packet, it forwards correctly with prob. 1/3 where the packet takes 2 seconds to reach the target; forwards it to a wrong path with prob. 1/3, where the packet comes back after 3 seconds; forwards it to another wrong with prob. 1/3, where the packet comes back after 5 seconds.

- Q: What is the expected time delay for the packet reach the target?
  - Memoryless
  - Expectation by condition

Example 2

- Suppose a spam filter gives each incoming email an overall score. A higher score means the email is more likely to be spam. By running the filter on training set of email (known normal + known spam), we know that 80% of normal emails have scores of 1.5 ± 0.4; 68% of spam emails have scores of 4 ± 1. Assume the score of normal or spam email follows normal distr.
- Q1: If we want spam detection rate of 95%, what threshold should we configure the filter?
- Q2: What is the false positive rate under this configuration?

- Terminology:
  - False positive: mistakenly treat normal event as abnormal event.
  - False negative: mistakenly treat abnormal event as normal event.

Example 3

- A ball is drawn from an bottle containing three white and two black balls. After each ball is drawn, it is then placed back. This goes on indefinitely.
  - Q: What is the probability that among the first four drawn balls, exactly two are white?

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

Example 4

- A type of battery has a lifetime with μ=40 hours and σ=20 hours. A battery is used until it fails, at which point it is replaced by a new one.
  - Q: If we have 25 batteries, what's the probability that over 1100 hours of use can be achieved?
  - Approximate by central limit theorem

Example 5

- If the prob. of a person suffer bad reaction from the injection of a given serum is 0.1%, determine the probability that out of 2000 individuals (a). exactly 3 (b). More than 2 individuals suffer a bad reaction? (c). If we inject one person per minute, what is the average time between two bad reaction injections?

  - Poisson distribution (for rare event in a large number of independent event series)
    - Can use Binomial, but too much computation
  - Geometric

Example 6

- A group of n camping people work on assembling their individual tent individually. The time for a person finishes is modeled by r.v. X.
  - Q1: what is the PDF for the time when the first tent is ready?
  - Q2: what is the PDF for the time when all tents are ready?

  - Suppose X_i are i.i.d., i=1, 2, $\cdots$, n
  - Q: compute PDF of r.v. Y and Z where
    - $Y = \max(X_1, X_2, \cdots, X_n)$
    - $Z = \min(X_1, X_2, \cdots, X_n)$
    - Y, Z can be used for modeling many phenomenon

Example 7

- A coin having probability p of coming up heads is flipped until two of the most recent three flips are heads. Let N denote the number of heads. Find E[N].

$$\underbrace{0\ 0\ 0\ 1}\ \underbrace{0\ 0\ 0\ 0\ 1}\ \underbrace{0\ 0\ 1}\ \underbrace{0\ 1}$$

- $P(N=n) = P(Y_1 \geq 3, \cdots, Y_{n-1} \geq 3, Y_n \leq 2)$