Geometrical Probability

Prof. Hai Jiang
Tsinghua University

Outline

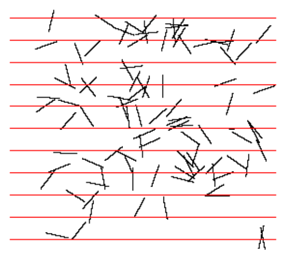
- Introduction to Geometrical Probability
- Functions of Random Variables
  - Many, many examples
- Perturbation Methods
  - Perturbation to RV
  - Perturbation to PDF
  - Perturbation to sample space
- Optimal facility location

Geometrical Probability

- It arises in problems dealing with distribution of objects in space
- In urban applications, geometrical probability concepts help to analyze relationships among objects distributed throughout an urban environment
  - People requiring on-scene service
    - Pickup by taxicab, on-scene medical care

Buffon’s Needle Experiment

- Parallel lines spaced at *d*
- Throw a needle of length l<d “at random”



- What’s the probability that this needle intersects one of the parallel lines

Functions of Random Variables

- Often we know the probability law of one or more RVs
- We wish to obtain the probability laws of other RVs expressed in terms of the original RVs
- This is called a problem of *derived distributions*, since we must derive the joint probability distributions for the RVs in the second set

The Cumulative Distribution Method for Deriving Distributions

- Original set: $\{X_1, X_2, \ldots, X_N\}$ with joint CDF $F_{X_1, X_2, \ldots X_N}()$
- Second set: $\{Y_1, Y_2, \ldots, Y_M\}$ and $Y_i=g_i(X_1, X_2, \ldots, X_N)$
- Calculate CDF

  $F_{Y_1, Y_2, \ldots Y_M}(y_1, y_2, \ldots, y_M)$

  $=P\{Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_M \le y_M\}$

  $=P\{Y_i=g_i(X_1, X_2, \ldots, X_N) \le y_i, i=1, 2, \ldots, M\}$
- For continuous RVs

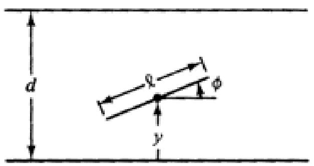
$$f_{Y_1,Y_2,\ldots,Y_M}(y_1,y_2,\ldots,y_M) = \frac{\partial^M}{\partial y_1 \partial y_2 \Lambda \, \partial y_M} F_{Y_1,Y_2,\ldots,Y_M}(y_1,y_2,\ldots,y_M)$$

- For discrete RVs

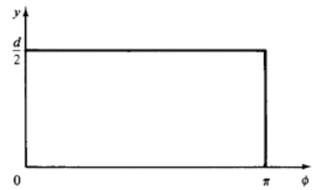
  Subtract appropriate CDF values

Buffon's Needle Experiment (Cont'd)

- Define RVs
  - Y: distance from the center of the needle to the closest of the parallel lines
  - Φ: angle of the needle



Joint Sample Space

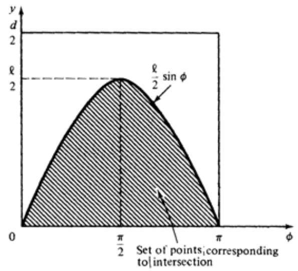


$$f_Y(y) = \frac{2}{d} \qquad 0 \le y \le \frac{d}{2}$$

$$f_\Phi(\phi) = \frac{1}{\pi} \qquad 0 \le \phi \le \pi$$

$$f_{Y,\Phi}(y,\phi) = f_Y(y) f_\Phi(\phi) = \frac{2}{d\pi} \qquad 0 \le y \le \frac{d}{2}, 0 \le \phi \le \pi$$

Work in the Joint Sample Space



In order to intersect for a given ϕ, we need

$$y \le \frac{l}{2}\sin\phi$$

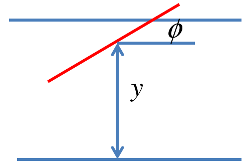
$$P=\int_0^{\pi}\int_0^{\frac{l}{2}\sin\phi}\frac{2}{\pi d}\,dy\,d\phi=\int_0^{\pi}\frac{l\sin\phi}{\pi d}\,d\phi=\frac{l}{\pi d}(-\cos\phi)\Big|_0^{\pi}=\frac{2l}{\pi d}$$

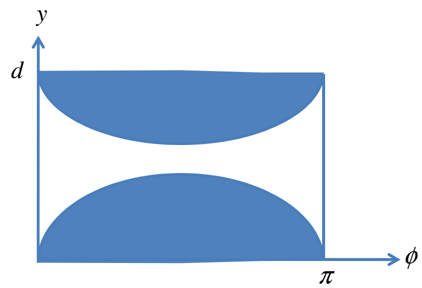
What can this result bring to us?

http://www.ms.uky.edu/~mai/java/stat/buff.html

Exercise

- Redo the analysis assuming that Y is the distance from the center of the needle to the next “southern” parallel line





$$\begin{cases} y \le \frac{l}{2}\sin\phi \\ d-y \le \frac{l}{2}\sin\phi \Rightarrow y \ge d-\frac{l}{2}\sin\phi \end{cases}$$

Response Distance of Police Car

- An police car patrols the highway back and forth to find motorists in need of assistance
- Also, accidents can occur that create a need for on scene assistance
- The police car is dispatched by radio to these accidents
- What’ s the probability law of the travel distance for the police car to reach a random accident

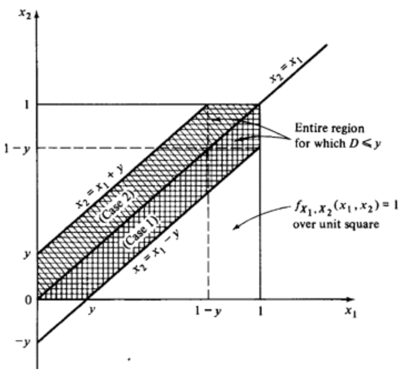
4-Step Process

- Define RVs
  - X_1=location of the accident
  - X_2=location of the police car at the moment of dispatch
- Joint sample space $0 \le X_1 \le 1$, $0 \le X_2 \le 1$
- Joint probability distribution

$$f_{X_1,X_2}(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2)=1\times1=1 \qquad 0\le x_1,x_2\le 1$$

- Work in the sample space

Probability Law of D=|X$_1$-X$_2$|



$$F_D(y) = P\{D \le y\} = P\{|x_1 - x_2| \le y\}$$
$$= 1 - 2\left[\frac{1}{2}(1-y)^2\right]$$
$$= 2y - y^2$$

$$f_D(y) = \begin{cases} 2(1-y) & y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

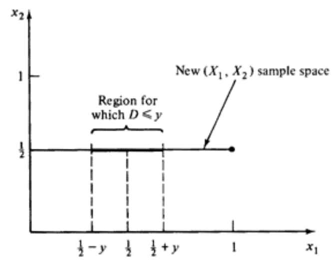
$$E[D] = \int_0^1 (2-2y)y\,dy = y^2 - \frac{2}{3}y^3\Big|_0^1 = \frac{1}{3}$$

$$E[D^2] = \int_0^1 (2-2y)y^2\,dy = \frac{2}{3}y^3 - \frac{1}{2}y^4\Big|_0^1 = \frac{1}{6}$$

$$Var(D) = E[D^2] - (E[D])^2 = \frac{1}{18}$$

A Different Deployment Strategy

- What about positioning the police car at the center of the highway?



$$D' = |x_1 - \tfrac{1}{2}|$$
$$F_{D'}(y) = P\{|x_1 - \tfrac{1}{2}| \le y\}$$
$$= P\{\tfrac{1}{2} - y \le x_1 \le \tfrac{1}{2} + y\}$$
$$= 2y \qquad y \in [0, \tfrac{1}{2}]$$

$$f_{D'}(y) = \begin{cases} 2 & y \in [0, \tfrac{1}{2}] \\ 0 & \text{otherwise} \end{cases}$$

$$E[D'] = \int_0^{1/2} 2y\,dy = y^2\Big|_0^{1/2} = \tfrac{1}{4}$$

$$E[D'^2] = \int_0^{1/2} 2y^2\,dy = \tfrac{2}{3}y^3\Big|_0^{1/2} = \tfrac{1}{12}$$

$$Var(D') = \tfrac{1}{12} - (\tfrac{1}{4})^2 = \tfrac{1}{48}$$

Scaling – Expectation and Variance

- After performing the analysis for a conveniently scaled problem, we often need to rescale it to suit the real-world situation
  - The length of highway may be 13 km
- Suppose *V*=*aW*+*b*, where *a* and *b* are known

$$E[V] = aE[W] + b$$
$$E[V^2] = E[(aW+b)^2] = a^2E[W^2] + 2abE[W] + b^2$$
$$Var(V) = E[V^2] - (E[V])^2$$
$$= a^2E[W^2] + 2abE[W] + b^2 - a^2(E[W])^2 - 2abE[W] - b^2$$
$$= a^2(E[W^2] - (E[W])^2) = a^2Var(W)$$

Scaling - PDF

- The probability law for *V*

$$F_V(v) = P\{V \le v\} = P\{aW + b \le v\}$$
$$= \begin{cases} P\left\{W \le \dfrac{v-b}{a}\right\} = F_W\left(\dfrac{v-b}{a}\right) & a > 0 \\ P\left\{W \ge \dfrac{v-b}{a}\right\} = 1 - F_W\left(\dfrac{v-b}{a}\right) & a < 0 \end{cases}$$

$$f_V(v) = \frac{1}{|a|} f_W\left(\tfrac{v-b}{a}\right) \qquad \text{for any } a \ne 0$$

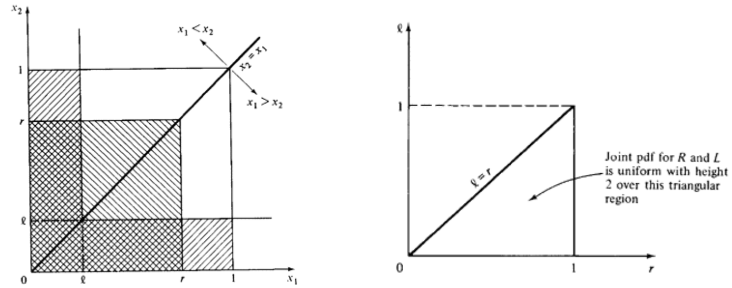
- Suppose the highway's length is 13 km, then D"=13D

$$f_{D''}(y) = \begin{cases} \dfrac{2}{13}\left(1 - \tfrac{y}{13}\right) & y \in [0,13] \\ 0 & \text{otherwise} \end{cases}$$

The Problem Revisited

- The road segment between the police car and the accident may be exposed to siren and lights
- We are interested in learning how this is going to disturb the neighborhood?
- Let L=min(X_1, X_2), R=max(X_1, X_2)
- What's the joint PDF for L and R?

Solution



$$F_{R,L}(r,l) = P\{R \le r, L \le l\}$$
$$= P\{\max(X_1, X_2) \le r \cap \min(X_1, X_2) \le l\}$$
$$= r^2 - (r-l)^2$$
$$= 2rl - l^2$$
$$f_{R,L}(r,l) = \begin{cases} 2 & 0 \le l \le r \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Expectations

$$f_L(l) = \int_l^1 f_{L,R}(l,r)dr = \int_l^1 2dr = 2(1-l) \qquad 0 \le l \le 1$$
$$E[L] = \int_0^1 l \cdot 2(1-l)dl = \left(l^2 - \frac{1}{3}l^3\right)\Big|_0^1 = \frac{1}{3}$$

$$f_R(r) = \int_0^r f_{L,R}(l,r)dl = \int_0^r 2dl = 2r \qquad 0 \le r \le 1$$
$$E[R] = \int_0^1 r \cdot 2rdl = \frac{2}{3}r^3\Big|_0^1 = \frac{2}{3}$$

What about Travel Time?

- Suppose we are interested in travel time required for the police car to reach the location of the accident
- Define RVs
  - T=travel time
  - X=travel distance
  - S=travel speed (independent of D)

$$S = \begin{cases} 1, \text{ w.p. } 1/2 \\ 2, \text{ w.p. } 1/2 \end{cases}$$

Unit Impulse Function

- Define Unite Impulse Function $\mu_0(x)$ has a area equal to 1 at x=0, that is,

$$\int_{-a}^{b} \mu_0(x)dx = 1 \qquad \forall a,b>0$$

- Bernoulli RV can be expressed as

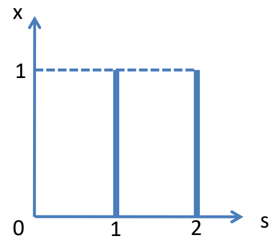
$$f_X(x) = (1-p)\mu_0(x) + p\mu_0(x-1)$$

Joint Sample Space for D and S

$$f_S(s) = \frac{1}{2}\mu_0(s-1) + \frac{1}{2}\mu_0(s-2), \qquad -\infty \le s \le +\infty$$

$$f_X(x) = 2(1-x) \qquad x \in [0,1]$$

$$f_{X,S}(x,s) = f_X(x) f_S(s) = (1-x)\mu_0(s-1) + (1-x)\mu_0(s-2) \qquad x \in [0,1], -\infty \le s \le +\infty$$



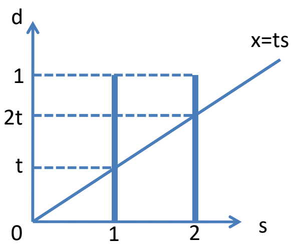
Next step: Work within the joint sample space to find $F_T(t)$

$$T = \frac{X}{S}$$

$$F_T(t) = P\left\{\frac{X}{S} \le t\right\}$$

Find $F_T(t)$

1. When $0 \le t \le 1/2$

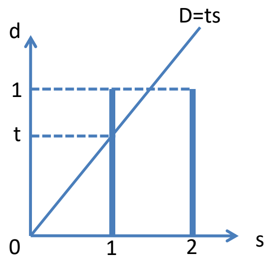


$$F_T(t) = \int_0^2 \int_0^{ts} (1\text{-}x)\mu_0(s\text{-}1) + (1\text{-}x)\mu_0(s\text{-}2)dxds$$

$$= \int_0^t (1-x)dx + \int_0^{2t} (1-x)dx = 3t - \frac{5}{2}t^2$$

$$f_T(t) = 3 - 5t$$

2. When $1/2 \le t \le 1$

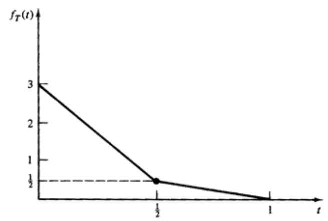


$$F_T(t) = \int_0^t (1-x)dx + \frac{1}{2} = t - \frac{t^2}{2} + \frac{1}{2}$$

$$f_T(t) = 1 - t$$

$f_T(t)$ and E[T]

$$f_T(t) = \begin{cases} 3-5t & 0 \le t \le \frac{1}{2} \\ 1-t & \frac{1}{2} \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

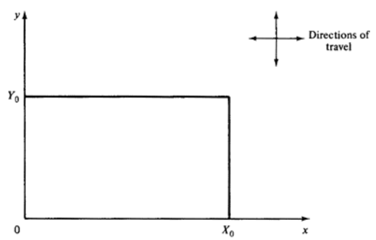


$$E[T] = \int_0^{1/2} t(3-5t)dt + \int_{1/2}^{1} t(1-t)dt = 1/4$$

A different approach yield an answer that is about 11% less

$$E[T] = E\left[\frac{D}{S}\right] = \frac{E[D]}{E[S]} = \frac{1/3}{3/2} = \frac{2}{9}$$

A Twist – Rectangular Response Area



- X_0-by-Y_0 rectangular response area
- Location of accident (X_1, Y_1) and of the police car (X_2, Y_2) are independently uniformly distributed over the area
- Travel distance occurs according to the "right-angle" metric
- What is expectation and variance of travel distance?

$$D = |X_1 - X_2| + |Y_1 - Y_2|$$

$$E[D] = E[|X_1 - X_2|] + E[|Y_1 - Y_2|] = \frac{1}{3}(X_0 + Y_0)$$

$$Var(D) = Var(|X_1 - X_2|) + Var(|Y_1 - Y_2|) = \frac{1}{18}(X_0^2 + Y_0^2)$$

Accuracy of AVL Systems

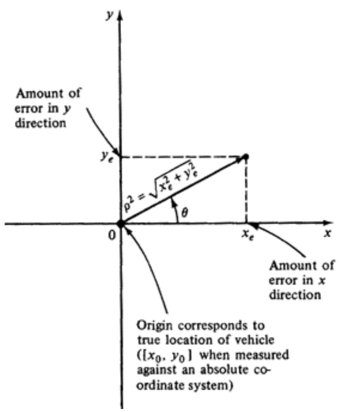
- Automatic Vehicle Location (AVL) systems are often installed to estimate the location of vehicles
- Suppose the vehicle is located at (X_0, Y_0) and AVL estimates the location to be
  - $X=X_0+X_e$
  - $Y=Y_0+Y_e$
  - X_e, Y_e are independent zero mean normal RVs

$$f_{X_e}(x) = f_{Y_e}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

- Derive the PDF for "radius of error"

$$R = \sqrt{X_e^2 + Y_e^2}$$

4-Step Process



$$1.\ \text{RVs}: X_e, Y_e$$

$$2.\ \text{Joint sample space}: -\infty \le X_e, Y_e \le +\infty$$

$$3.\ \text{PDF}: f_{X_e,Y_e}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad -\infty \le X_e, Y_e \le +\infty$$

$$4.\ F_R(r) = P\{R \le r\} = P\left\{\sqrt{X_e^2 + Y_e^2} \le r\right\}$$

$$= \iint_{\substack{\text{circle of}\\ \text{radius } r}} \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} dxdy = \int_0^{2\pi}\int_0^r \frac{1}{2\pi\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} \rho d\rho d\theta$$

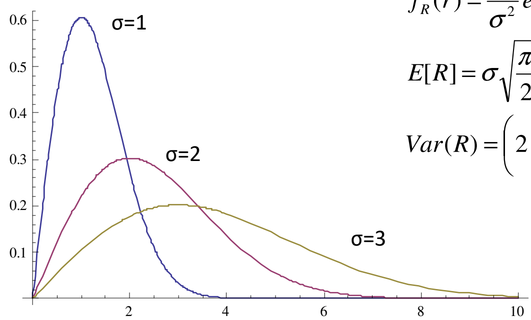
$$= 2\pi\int_0^r \frac{1}{2\pi\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} \rho d\rho = -\int_0^r e^{-\frac{\rho^2}{2\sigma^2}} d\left(-\frac{\rho^2}{2\sigma^2}\right)$$

$$= \left(-\exp\left(\frac{\rho^2}{2\sigma^2}\right)\right)\Bigg|_0^r = 1 - e^{-\frac{r^2}{2\sigma^2}}$$

$$5.\ f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \qquad r \ge 0$$

Rayleigh PDf

Rayleigh PDF with parameter σ

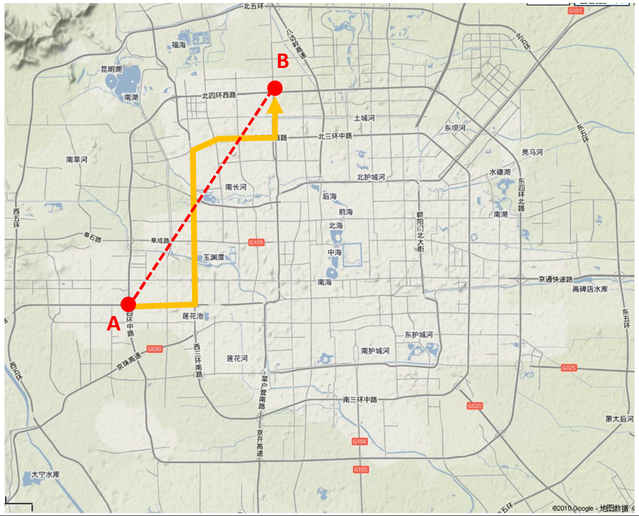


$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \qquad r \ge 0$$

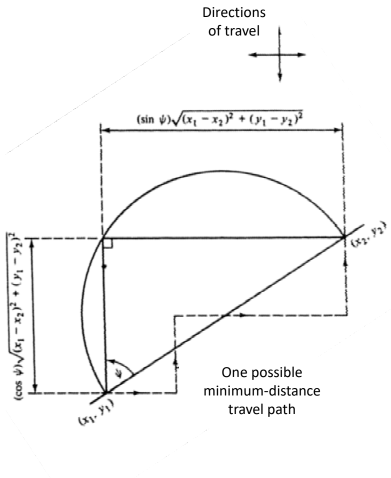
$$E[R] = \sigma\sqrt{\frac{\pi}{2}} \approx 1.25\sigma$$

$$Var(R) = \left(2 - \frac{\pi}{2}\right)\sigma^2 \approx (0.65\sigma)^2$$

Ratio of Right Angle to Euclidean Distance Metrics



Ratio of the Right Angle and Euclidean Distance



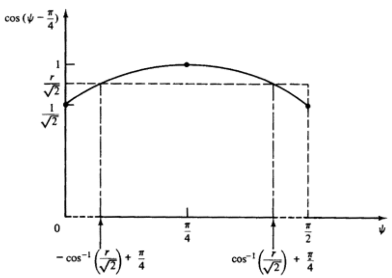
- Let ψ be the angle at which the directions of travel are rotated w.r.t. the straight line connecting the two points
- In a large, uniform city it make sense to assume that ψ is uniformly distributed between 0 and $\pi/2$
- The right angle distance is

$$(\cos\psi+\sin\psi)\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2}$$

- The ratio

$$R=\cos\psi+\sin\psi=\sqrt{2}\cos(\psi-\frac{\pi}{4})$$

Working in the Sample Space



$$F_R(r)=P\left\{\sqrt{2}\cos(\psi-\frac{\pi}{4})\le r\right\}$$

$$=P\left\{\cos(\psi-\frac{\pi}{4})\le\frac{r}{\sqrt{2}}\right\}$$

$$=\left(-\cos^{-1}\left(\frac{r}{\sqrt{2}}\right)+\frac{\pi}{4}\right)\Big/\frac{\pi}{4}$$

$$=1-\frac{4}{\pi}\cos^{-1}\left(\frac{r}{\sqrt{2}}\right)\qquad r\in[1,\sqrt{2}]$$

$$f_R(r)=-\frac{4}{\pi}\left(-\frac{1}{\sqrt{1-\left(\frac{r}{\sqrt{2}}\right)^2}}\frac{1}{\sqrt{2}}\right)=\frac{4}{\pi}\frac{1}{\sqrt{2-r^2}}$$

$$E[R]=\frac{4}{\pi}\approx 1.273$$

$$Var(R)=1+\frac{2}{\pi}-\frac{16}{\pi^2}\approx 0.0155$$

A “robust” estimate: the vehicle travels about 1.273 times the Euclidean distance. (Coefficient of Variation = 0.098)

Final Example of Derived Distributions

- Assume that we are running an experiment to estimate the distribution of distance travelled by police cars
- All we have available are travel distances recorded as 0 miles, 1 mile, 2 miles, and so on
- Before the police car is dispatched, the policeman reads the integer part of the odometer. At the end of the trip, he reads the integer part of the odometer again and records the difference in miles
- These differences are recorded in our data

Example

	Odometer	Recorded mileage
Prior to dispatch	5.9	5
Actual distance	1.2	
After dispatch	7.1	7
Distance recorded		7-5=2

- If the vehicle had traveled 0.6 mile since the last reading, the recorded mileage would be 1 mile
- What is the effect of such truncation?

4-Step Process

- Define RVs

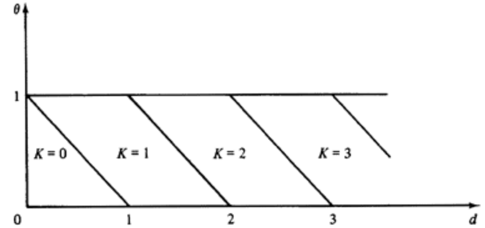
  D : actual distance traveled

  Θ : accumulated noninteger odometer mileage at the moment of dispatch (uniform between 0 and 1). D and Θ are independent

  K : recorded mileage reading

  $$K = \lfloor D + \Theta \rfloor$$

- Joint sample space



4-Step Process

- Joint PDF

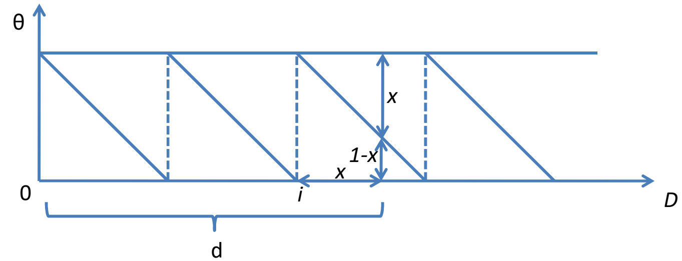
$$f_{D,\Theta}(d,\theta) = f_D(d) f_\Theta(\theta) = f_D(d) \quad d \ge 0, 0 \le \theta \le 1$$

- Work in the joint sample space
  - If the cdf of D is known

$$P\{K = k\} = \int_0^1 \int_{k-\theta}^{k-\theta+1} f_D(x)\,dx\,d\theta = \int_0^1 [F_D(k-\theta+1) - F_D(k-\theta)]\,d\theta$$

Interesting Results about K and D

- E[K] and E[D]



$$E[K \mid D = d] = E[K \mid D = i + x] = i(1-x) + (i+1)x = i + x = d$$

$$E[K] = \sum_{d=0}^{\infty} P(D=d) E[K \mid D=d] = \sum_{d=0}^{\infty} P(D=d) d = E[D]$$

Regardless of the functional form of $F_D()$, E[K] is an unbiased estimate of E[D].
Note: Zero-mileage trips have to be recorded and used.

Interesting Results about K and D

- Var(K) and Var(D)

$$E[K^2 \mid D=d] = E[K^2 \mid D=i+x] = i^2(1-x)+(i+1)^2 x$$
$$= (i+x)^2 - x^2 + x = d^2 + (x - x^2)$$

$$E[K^2] = \sum_{d=0}^{\infty} P(D=d)E[K^2 \mid D=d]$$
$$= \sum_{d=0}^{\infty} P(D=d)\left[d^2 + (x-x^2)\right]$$
$$= E[D^2] + \Delta \qquad \Delta \geq 0$$

$$Var(K) = E[K^2] - (E[K])^2 = E[D^2] + \Delta - (E[D])^2 = Var(D) + \Delta \geq Var(D)$$

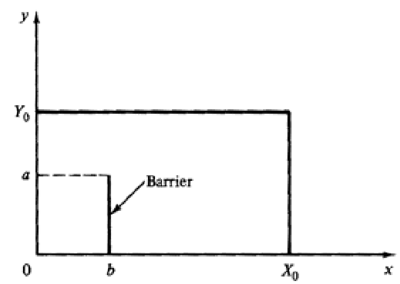
Regardless of the functional form of FD(), Var(K) overestimates Var(D).

Perturbation Method

Motivation

- The probabilistic behavior of RVs may be complicated by realistic urban environment
  - Barriers to travel (e.g., parks, rivers)
  - Irregular area topology
- Perturbation Methods
  - Solve a simpler problem by ignoring the complication
  - Add the complication as perturbation to the simpler problem
- We discuss three situations
  - Perturbation to the random variable
  - Perturbation to a PDF
  - Geometrical perturbation to the region over which objects are uniformly distributed

Perturbation to RV: Barriers to Travel



- The incident's position (X_1,Y_1) and response unit's position (X_2,Y_2) are uniformly independently distributed. Travel has to be right-angle
- We want to know how the barrier increases travel distance

When There is No Barrier

- If there isn't such a barrier

$$E[D] = \tfrac{1}{3}(X_0 + Y_0)$$

- With the barrier, the new travel distance can be expressed as the sum of the "old" travel distance and a perturbation distance $D' = D + D_e$

$$E[D]' = E[D] + E[D_e]$$

Calculate the Perturbation

$$E[D_e] = P\{A_1\}E[D_e \mid A_1] + P\{A_2\}E[D_e \mid A_2]$$

A_1 : scenarios where $D_e > 0$

A_2 : scenarios where $D_e = 0$

Event A_1 requires that the response unit and the customer be on opposite side of the barrier, that is,

$$Y_1 < a, Y_2 < a, \min(X_1, X_2) \le b, \max(X_1, X_2) > b$$

$$P\{A_1\} = 2\left(\tfrac{b}{X_0}\tfrac{a}{Y_0}\right)\left(\tfrac{X_0-b}{X_0}\tfrac{a}{Y_0}\right)$$

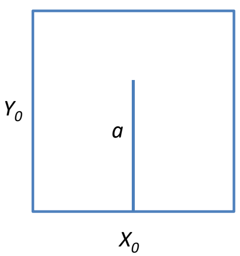
Given A_1, the extra distance travelled is

$$\{D_e \mid A_1\} = 2\min\left(\underbrace{a - \{Y_1 \mid A_1\}}_{Z_1}, \underbrace{a - \{Y_2 \mid A_1\}}_{Z_2}\right)$$

Since Z_1, Z_2 are uniformly independently distributed on [0,a], we have $E[D_e \mid A_1] = 2 \times \frac{a}{3} = \frac{2a}{3}$

Results

We therefore have $E[D'] = \tfrac{1}{3}(X_0 + Y_0) + \tfrac{4a^3 b}{3X_0^2 Y_0^2}(X_0 - b)$

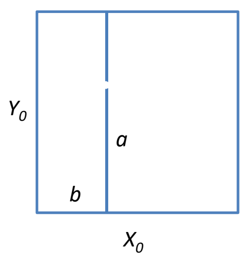


Suppose $X_0 = Y_0 = 1 \quad b = \tfrac{1}{2}$

$$E[D'] = \frac{2}{3} + \frac{a^3}{3}$$

If $a = \frac{1}{2}$, extra distance traveled is $\frac{a^3}{2} = 6.25\%$ more

A Complete Barrier?



The extra distance travelled:

$$E[D_e] = P\{A_{1a}\}E[D_e \mid A_{1a}] + P\{A_{1b}\}E[D_e \mid A_{1b}] + P\{A_2\}E[D_e \mid A_2]$$

A_{1a} : scenario where $D_e > 0$ and $Y_1, Y_2 > a$

A_{1b} : scenario where $D_e > 0$ and $Y_1, Y_2 \le a$

A_2 : scenario where $D_e = 0$

$$E[D_e] = 2\left(\frac{b}{X_0}\frac{Y_0 - a}{Y_0}\right)\left(\frac{X_0 - b}{X_0}\frac{Y_0 - a}{Y_0}\right)\cdot\frac{2}{3}(Y_0 - a) + 2\left(\frac{b}{X_0}\frac{a}{Y_0}\right)\left(\frac{X_0 - b}{X_0}\frac{a}{Y_0}\right)\cdot\frac{2a}{3}$$

$$= \frac{4b(X_0 - b)}{3X_0^2 Y_0^2}\left[a^3 + (Y_0 - a)^3\right]$$

Perturbations to a PDF

- Suppose $f_{X'}$ (x) is the pdf of interest, f_X(x) is a simpler pdf, h(x) is an added and g(x) a subtracted perturbation term

$$f_{X'}(x) = f_X(x) - g(x) + h(x)$$

where $\quad f_X(x) - g(x) + h(x) \geq 0,\ g(x), h(x) \geq 0$

$$\int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty} h(x)dx = P_\Delta$$

$$\frac{g(x)}{P_\Delta} \rightarrow X_g$$

$$\frac{h(x)}{P_\Delta} \rightarrow X_h$$

Expectation and Variance of X′

$$E[X'] = \int_{-\infty}^{\infty} x\left[f_X(x) - g(x) + h(x)\right]dx$$

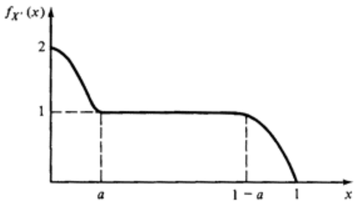
$$= \int_{-\infty}^{\infty} x f_X(x)dx - P_\Delta \int_{-\infty}^{\infty} x\frac{g(x)}{P_\Delta}dx + P_\Delta \int_{-\infty}^{\infty} x\frac{h(x)}{P_\Delta}dx$$

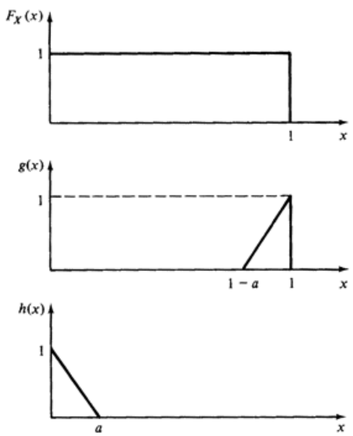
$$= E[X] - P_\Delta\left(E[X_g] - E[X_h]\right)$$

$$E[X'^2] = E[X^2] - P_\Delta\left(E[X_g^2] - E[X_h^2]\right)$$

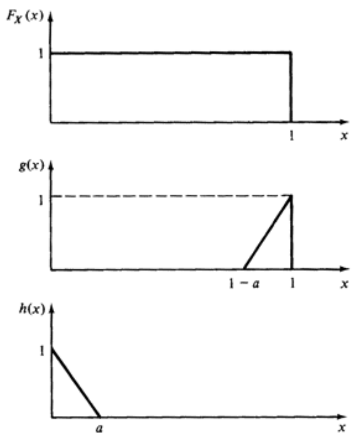
$$Var(X') = E[X'^2] - E[X']^2$$

Example: Triangle Perturbation





Calculation of E[X] and Var(X′)



$$E[X] = \frac{1}{2} \qquad \sigma_X^2 = \frac{1}{12}$$

$$E[X_g] = 1 - \frac{a}{3} \qquad \sigma_{X_g}^2 = \frac{a^2}{18} \qquad E[X_g^2] = \frac{a^2}{18} + \left(1 - \frac{a}{3}\right)^2$$

$$E[X_h] = \frac{a}{3} \qquad \sigma_{X_h}^2 = \frac{a^2}{18} \qquad E[X_h^2] = \frac{a^2}{6}$$

$$E[X'] = \frac{1}{2} - \frac{a}{2}\left(1 - \frac{a}{3} - \frac{a}{3}\right) = \frac{1}{2} - \frac{a}{2}\left(\frac{3 - 2a}{3}\right)$$

$$= \frac{1}{2}\left[1 - \frac{1}{3}a(3 - 2a)\right]$$

$$\sigma_{X'}^2 = \frac{1}{12} - \frac{a}{2}\left[\frac{a^2}{18} + \left(1 - \frac{a}{3}\right)^2 - \frac{a^2}{6}\right] + 2\frac{a}{2}\frac{1}{2}\left(1 - \frac{a}{3} - \frac{a}{3}\right)$$

$$-\frac{a^2}{4}\left(1 - \frac{a}{3} - \frac{a}{3}\right)^2$$

Do the results make sense when a=1/2?

Perturbations to a Sample Space

- Suppose that an object with location (X,Y) is uniformly distributed over S
- Now we perturb the sample space by the addition and subtraction of a term

$$S' = S - S_\Delta^1 + S_\Delta^2$$

where S_Δ^1 and S_Δ^2 have equal area

- For function g(X,Y) over S', we have

$$E_{S'}[g(X,Y)] = E_S[g(X,Y)] - P_\Delta\left(E_{S_\Delta^1}[g(X,Y)] - E_{S_\Delta^2}[g(X,Y)]\right)$$

$$E_{S'}[g^2(X,Y)] = E_S[g^2(X,Y)] - P_\Delta\left(E_{S_\Delta^1}[g^2(X,Y)] - E_{S_\Delta^2}[g^2(X,Y)]\right)$$

$$\text{where } P_\Delta = \frac{\text{area of } S_\Delta^1}{\text{area of } S}$$

$$E_{S'}[g(X,Y)] = \iint_{(x,y)\in S'} g(X,Y)\frac{1}{\text{area of } S}dxdy$$

$$= \iint_{(x,y)\in S} g(X,Y)\frac{1}{\text{area of } S}dxdy - \iint_{(x,y)\in S_\Delta^1} g(X,Y)\frac{1}{\text{area of } S}dxdy$$

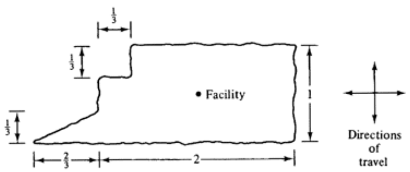
$$+ \iint_{(x,y)\in S_\Delta^2} g(X,Y)\frac{1}{\text{area of } S}dxdy$$

$$= \iint_{(x,y)\in S} g(X,Y)\frac{1}{\text{area of } S}dxdy - \frac{\text{area of } S_\Delta^1}{\text{area of } S}\iint_{(x,y)\in S_\Delta^1} g(X,Y)\frac{1}{\text{area of } S_\Delta^1}dxdy$$

$$+ \frac{\text{area of } S_\Delta^1}{\text{area of } S}\iint_{(x,y)\in S_\Delta^2} g(X,Y)\frac{1}{\text{area of } S_\Delta^2}dxdy$$

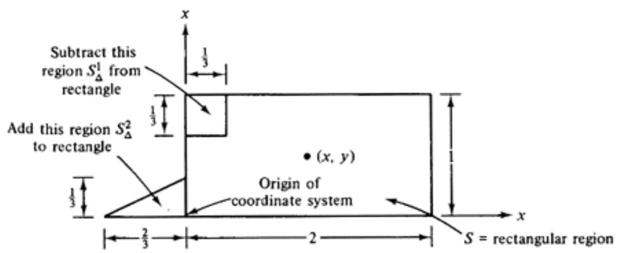
$$= E_S[g(X,Y)] - \frac{\text{area of } S_\Delta^1}{\text{area of } S}\left(E_{S_\Delta^1}[g(X,Y)] - E_{S_\Delta^2}[g(X,Y)]\right)$$

Facility Location in an Irregular District

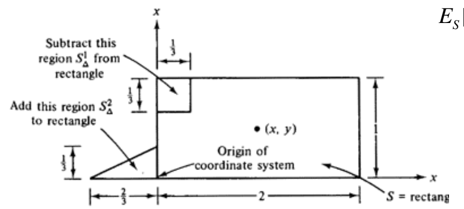


- Service requests are iid uniform
- Assume that travel is right-angle
- Determine the location of the single facility (x_0,y_0) to minimize the expected travel distance E[g(X,Y)]

$$g(X,Y) = |X - x_0| + |Y - y_0|$$



Facility Location in an Irregular District



$$E_S[g(X,Y)] = E[|X - x_0|] + E[|Y - y_0|]$$

$$= P\{X \le x_0\}\tfrac{1}{2}x_0 + P\{X > x_0\}\tfrac{1}{2}(2 - x_0)$$

$$+ P\{Y \le y_0\}\tfrac{1}{2}y_0 + P\{Y > y_0\}\tfrac{1}{2}(1 - y_0)$$

$$= \tfrac{x_0^2}{4} + \tfrac{(2-x_0)^2}{4} + \tfrac{y_0^2}{2} + \tfrac{(1-y_0)^2}{2}$$

$$P_\Delta = \frac{\frac{1}{3}\frac{1}{3}}{2} = \tfrac{1}{18}$$

$$E_{S_\Delta^1}[g(X,Y)] = (x_0 - \tfrac{1}{6}) + (1 - y_0 - \tfrac{1}{6})$$

$$E_{S_\Delta^2}[g(X,Y)] = (x_0 + \tfrac{1}{3}\tfrac{2}{3}) + (y_0 - \tfrac{1}{3}\cdot\tfrac{1}{3})$$

$$E_{S'}[g(X,Y)] = \tfrac{x_0^2}{4} + \tfrac{(2-x_0)^2}{4} + \tfrac{y_0^2}{2} + \tfrac{(1-y_0)^2}{2} - \tfrac{1}{18}\left(\tfrac{5}{9} - 2y_0\right)$$

$$\frac{\partial E_S[g(X,Y)]}{\partial x_0} = 0 \Rightarrow x_0 = 1$$

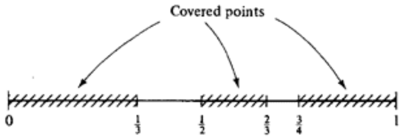
$$\frac{\partial E_S[g(X,Y)]}{\partial y_0} = 0 \Rightarrow y_0 = \tfrac{4}{9}$$

Coverage Problem

- In deployment applications, we may be interested in knowing the fraction of city which receives "adequate coverage" by the service
- Coverage is often defined as travel time being less than or equal to a threshold
  - The Emergency Medical Services Act of 1973 stipulates that 95% of ambulance responses should occur in less than 30 minutes
- Coverage problems are defined as calculating the probabilities that certain fixed geometrical figures or sets in the plane are covered by other figures whose position is in some way random

Coverage by Randomly Positioned Ambulances

- X = set of points in [0,1] which are covered
- μ(X) = "length" of the set X



$$\mu(X) = \tfrac{1}{3} + \tfrac{1}{6} + \tfrac{1}{4} = \tfrac{3}{4}$$

- A point is said to be covered if at least one ambulance is within a distance of ½a
- We have N ambulances distributed iid uniformly over [-½a,1+½a]

Solution

The probability that any point *x* is covered by a particular ambulance:

$$p_1(x) = \frac{a}{1+a} = p_1$$

The probability that point *x* is not covered by any ambulance:

$$(1-p_1)^N$$

The probability that point *x* is covered by at least one ambulance:

$$p_N(x) = 1-(1-p_1)^N = p_N$$

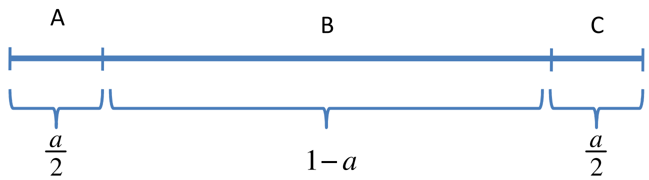
Define indicator variable:

$$S(x) = \begin{cases} 1 & \text{if } x \text{ is covered} \\ 0 & \text{otherwise} \end{cases}$$

$$E[\mu(X)] = E[\int_0^1 S(x)dx] = \int_0^1 E[S(x)]dx = \int_0^1 p_N dx = p_N = 1-\left(1-\frac{a}{1+a}\right)^N$$

Boundary Effects

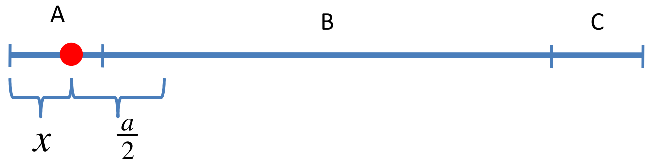
- Suppose that the ambulances are distributed over [0,1], what is E[μ(X)]?



$$\begin{aligned} E[\mu(X)] &= E[\mu_A(X)+\mu_B(X)+\mu_C(X)] \\ &= 2E[\mu_A(X)] + E[\mu_B(X)] \end{aligned}$$

$$E[\mu_B(X)] = (1-a)\left[1-(1-a)^N\right]$$

Boundary Effects (Cont'd)



Probability for *x* to be covered by a particular ambulance is $\frac{a}{2} + x$

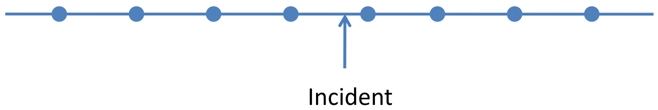
Probability for *x* to be covered by any ambulance is $1-(1-\frac{a}{2}-x)^N$

$$E[\mu_A(X)] = E[\int_0^{a/2} S(x)dx] = \int_0^{a/2} E[S(x)]dx = \int_0^{a/2}\left(1-(1-\tfrac{a}{2}-x)^N\right)dx$$

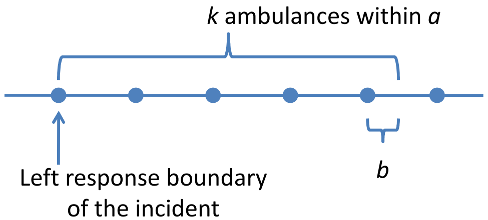
$$E[\mu(X)] = (1-a)\left[1-(1-a)^N\right] + 2\int_0^{a/2}\left(1-(1-\tfrac{a}{2}-x)^N\right)dx$$

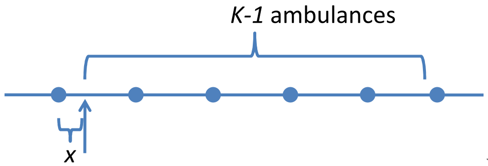
Coverage by Fixed-Position Ambulances

- We can also position the response units at fixed positions rather than random
- Suppose they are positioned at unit separation
- We want to know the expected number of ambulances within a distance of a/2 of the random incident



Solution





x is uniform over [0,1]

Probability	# of Ambulances
1-b	k-1
b	k

$$E[N] = bk + (1-b)(k-1) = k-1+b = a$$

$$Var(N) = E[N^2] - (E[N])^2 = (k-1)^2(1-b) + k^2 b - (k-1+b)^2 = b(1-b)$$

Optimal District Design Problem

- How to design the shape of a response district given
  - Locations of a medical emergency (X_1, Y_1) and of the ambulance (X_2,Y_2) are iid uniform
  - Travel is parallel to the sides of the rectangular response district
- minimize $E[D]=(X_0+Y_0)/3$
- Subject to $X_0Y_0=A_0$

Optimal District Design Problem

$$\min\ E[D] = \tfrac{1}{3}\left(X_0 + Y_0\right)$$
$$\text{st}\quad X_0 Y_0 = A_0$$
$$X_0 = Y_0 = \sqrt{A_0}$$
$$E[D] = \tfrac{2}{3}\sqrt{A_0}$$

$$\min\ E[T] = \tfrac{1}{3}\left(\tfrac{X_0}{v_x} + \tfrac{Y_0}{v_y}\right)$$
$$\text{st}\quad X_0 Y_0 = A_0$$
$$\frac{X_0}{v_x} = \frac{Y_0}{v_y} = \sqrt{\frac{A_0}{v_x v_y}}$$
$$E[T] = \tfrac{2}{3}\sqrt{\tfrac{A_0}{v_x v_y}}$$

These estimates are pretty robust

Assume that $X_0 = \alpha Y_0$

Since $X_0 Y_0 = A_0$, we have $X_0 = \sqrt{A_0 \alpha}, Y_0 = \sqrt{\tfrac{A_0}{\alpha}}$

$$E[D] = \tfrac{1}{3}\left(\sqrt{\alpha A_0} + \sqrt{\tfrac{A_0}{\alpha}}\right) = \tfrac{2}{3}\sqrt{A_0} + \frac{\left(\sqrt{\alpha} - 1\right)^2}{3\sqrt{\alpha}}\sqrt{A_0}$$

When α=1.5, we would underestimate by about 2%

Spatial Poisson Process

- Poisson process in time

$$P(X(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \qquad t \ge 0, k = 0,1,2,\ldots$$

- Spatial Poisson process
  - Suppose entities are distributed around the city in a completely random manner
  - SPP describes the numbers of entities in a given area
  - λ is the intensity parameter

$$P(X(S) = k) = \frac{(\lambda A(S))^k e^{-\lambda A(S)}}{k!} \qquad A(S) \ge 0, k = 1,2,\ldots,$$

Example

- Suppose that emergency response units are distributed in a large region as a two-dimensional Poisson process with intensity parameter λ units per square mile
- What's the PDF of the travel distance between an incident and the nearest response unit?
- Assume Euclidean travel distance

Solution (1)

- Assume the incident occurs at some arbitrary point (x,y)
- Construct a circle of radius d centered at (x,y)
- The probability that there are no response units within the circle is

$$P(X(S) = 0) = \frac{\left(\lambda \pi d^2\right)^0 e^{-\lambda \pi d^2}}{0!} = e^{-\lambda \pi d^2}$$

- CDF

$$F_D(d) = P\{\text{at least one response unit in the circle}\} = 1 - e^{-\lambda \pi d^2}$$

Solution (2)

$$f_D(d) = 2d\lambda\pi e^{-\lambda\pi d^2} = \frac{d}{\left(\frac{1}{\sqrt{2\lambda\pi}}\right)^2}\exp\left(-\frac{d^2}{2\left(\frac{1}{\sqrt{2\lambda\pi}}\right)^2}\right) = \frac{d}{\sigma^2}\exp\left(-\frac{d^2}{2\sigma^2}\right)$$

This is Rayleigh distribution with parameter $\sigma = \frac{1}{\sqrt{2\lambda\pi}}$

$$E[D] = \sigma\sqrt{\frac{\pi}{2}} = \frac{1}{\sqrt{2\pi\lambda}}\sqrt{\frac{\pi}{2}} = \frac{1}{2\sqrt{\lambda}} = \frac{0.5}{\sqrt{\lambda}}$$

$$Var(D) = \left(2-\frac{\pi}{2}\right)\sigma^2 = \left(2-\frac{\pi}{2}\right)\frac{1}{2\pi\lambda} = \frac{0.068}{\lambda}$$

Example (cont' d)

- What if travel is right-angle?

$$P(X(S)=0) = \frac{\left(2\lambda d^2\right)^0 e^{-\lambda 2d^2}}{0!} = e^{-\lambda 2d^2}$$

$$F_D(d) = P\{\text{at least one response unit in the diamond}\} = 1 - e^{-2\lambda d^2}$$

$$f_D(d) = 4d\lambda e^{-\lambda 2d^2} = \frac{d}{\left(\frac{1}{\sqrt{4\lambda}}\right)^2}\exp\left(-\frac{d^2}{2\left(\frac{1}{\sqrt{4\lambda}}\right)^2}\right) = \frac{d}{\sigma^2}\exp\left(-\frac{d^2}{2\sigma^2}\right)$$

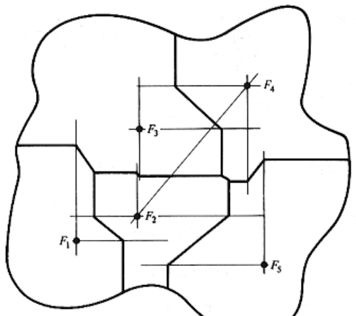
$$E[D] = \sigma\sqrt{\frac{\pi}{2}} = \frac{1}{\sqrt{4\lambda}}\sqrt{\frac{\pi}{2}} = \sqrt{\frac{\pi}{8\lambda}} = \frac{0.627}{\sqrt{\lambda}}$$

$$Var(D) = \left(2-\frac{\pi}{2}\right)\sigma^2 = \left(2-\frac{\pi}{2}\right)\frac{1}{4\lambda} = \frac{0.1075}{\lambda}$$

Facility Location and Districting

- Spatial Poisson Process can be applied to the problem of facility location and districting of a city
- Suppose demands are distributed uniformly throughout the plane and travel is right angle
  - For each service request, dispatch a response unit from the nearest facility
  - The individual requiring service travels to the nearest facility
- Each district around a facility would consist of all points closer to that facility than to any other

Illustrative Districting



F_i = location of facility i

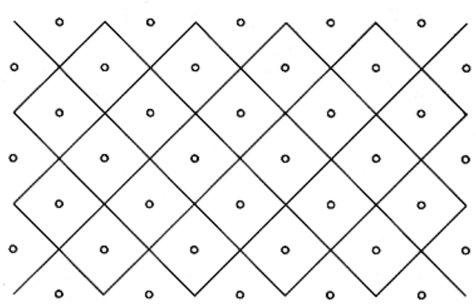
We want to investigate the mean travel distance for different facility location strategies

Upper Bound on Mean Travel Distance

- If we assume that facilities are distributed as a homogeneous spatial Poisson process, this corresponds to a totally unplanned system
- It is equivalent to throwing darts at a map of the city while closing your eyes
- We know the mean travel distance is

$$E[D] = \sqrt{\frac{\pi}{8\lambda}} \approx 0.627\frac{1}{\sqrt{\lambda}} \qquad \text{where } \lambda \text{ is the average density of facilities}$$

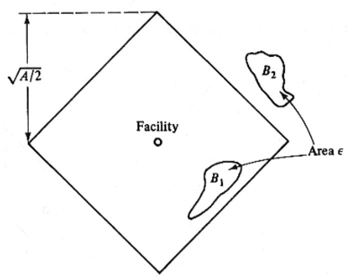
Lower Bound on Mean Travel Distance



- To achieve minimum mean travel distance, the facilities should be positioned in a regular lattice

Proof - Lemma 1

- Given that a facility's district must contain an area A, a square district rotated at 45°, centered at the facility's position, results in minimum mean travel distance



Proof: Suppose there is some redesign that results in lower mean travel distance.

$$E[D'] = E[D] - \tfrac{\varepsilon}{A}E[D \mid B_1] + \tfrac{\varepsilon}{A}E[D \mid B_2]$$
$$E[D'] - E[D] = \tfrac{\varepsilon}{A}\left(E[D \mid B_2] - \tfrac{\varepsilon}{A}E[D \mid B_1]\right)$$

However,

$$E[D \mid B_1] \le \sqrt{\tfrac{A}{2}},\ E[D \mid B_2] \ge \sqrt{\tfrac{A}{2}}$$

Therefore,

$$E[D'] - E[D] \ge 0, \text{ which is a contradiction!}$$

Proof - Lemma 2

- Given that
  - We have N square districts, each rotated at 45° and centered at the respective facility's position
  - The total area of the N districts equal NA
- We have that minimal mean travel time is obtained by setting the area of each district equal to A
- Suppose district i has an area of A_i, for a random service request

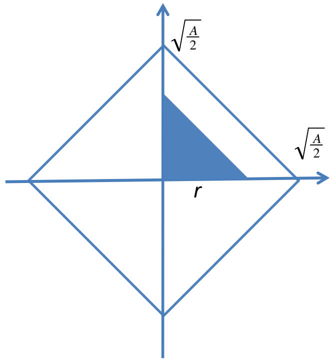
$$E[D] = \sum_{i=1}^{N} \tfrac{A_i}{NA}\tfrac{2}{3}\sqrt{\tfrac{A_i}{2}}$$

Within a Rotated Square District

$$F_D(r) = \frac{\frac{r^2}{2}}{\frac{A}{4}} = \frac{2r^2}{A}$$

$$f_D(r) = \frac{4r}{A}$$

$$E[D] = \int_0^{\sqrt{\frac{A}{2}}} r\frac{4r}{A}dr = \frac{2}{3}\sqrt{\frac{A}{2}}$$



Proof - Lemma 2

$$\min E[D] = \sum_{i=1}^{N} \frac{A_i}{NA}\frac{2}{3}\sqrt{\frac{A_i}{2}}$$

$$\text{st} \quad \sum_{i=1}^{N} A_i = NA$$

$$\Rightarrow A_i = A, \quad i = 1,2,\ldots,N$$

$$E[D] = \frac{2}{3}\sqrt{\frac{A}{2}} = \frac{2}{3}\sqrt{\frac{1}{2\lambda}} \approx 0.472\frac{1}{\sqrt{\lambda}}$$

Compared to our "completely unplanned" system, we can save mean travel distance by 25%

Summary

- Completed a tour of derived distributions, geometrical probability, and spatial Poisson process.
- Introduced the 4-Step process to calculate derived distributions
- Covered a wide range of applications in an urban setting