ECO239

Week 11

Poisson Distribution

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

- X: # of successes over a given time/space
- λ : expected number of successes per given time/space.

Should be the same unit!

Binomial Probability Distribution

$$P(X) = \frac{n!}{x!\,(n-x)!} P^{x}(1-P)^{n-x}$$

- P(X): probability of x successes in n trials, with P = probability of success.

- X: # of success in sample. (x= 0, 1, 2, …n)
- n: sample size
- P: probability of success.

Poisson Approximation to the Binomial Distribution

- When n is very large and p is small (np≤ 7), then we can **use Poisson distribution** to **approximate binomial distribution.**

- X: number of successes from n independent trials
- P: probability of success
- Distribution of X: binomial with mean np.
- n: large, np≤ 7.

- This binomial distribution can be approximated by the Poisson distribution with λ = np.

Poisson Approximation to Binomial Distribution ($np \leq 7$)

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

Practice

- 3.5% of small corporate would file for bankruptcy in the coming year. For a random sample of 100 small corporations, what is the probability that at least 3 will file for bankruptcy in the next year?

If you try to solve this problem using Binomial Distribution…

- $P(X \geq 3 | n=100, p=0.035) = 1 - P(X \leq 2 | n=100, p=0.035)$
- $= 1 - [P(X=0) + P(X=1) + P(X=2)]$
- $= 1 -$

$$\left[\frac{100!}{0!\,100!}(0.035^0)(0.965^{100}) + \frac{100!}{1!\,99!}(0.035^1)(0.965^{99}) + \frac{100!}{2!\,98!}(0.035^2)(0.965^{98})\right]$$

- $= 0.6841$

Practice

- 3.5% of small corporate would file for bankruptcy in the coming year. For a random sample of 100 small corporations, what is the probability that at least 3 will file for bankruptcy in the next year?

- Solve with Poisson Approximation to Binomial Distribution.

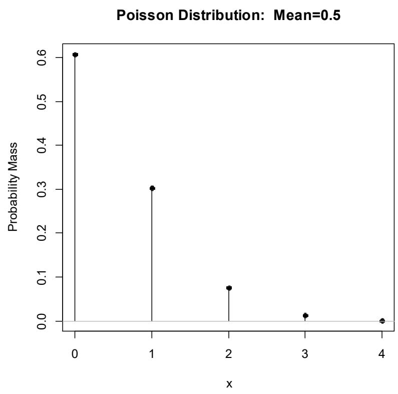
$$P(X) = \frac{e^{-\lambda}\lambda^x}{x!}$$

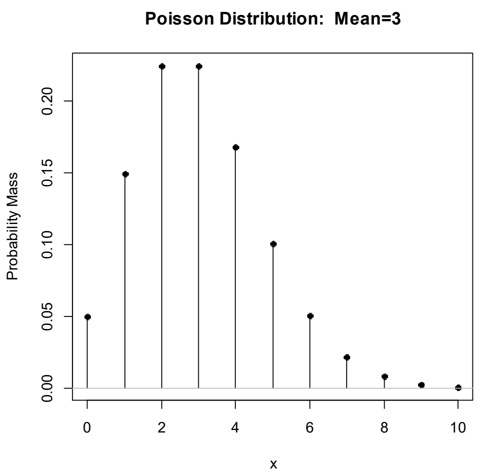
- n= 100, p = 0.035, np = 3.5
- $P(X \geq 3 | np=3.5) = 1 - P(X \leq 2 | np=3.5)$
- $= 1 - [P(X=0) + P(X=1) + P(X=2)]$
- =1- e^(-3.5)*[1+3.5+(3.5^2)/2]=0.684093

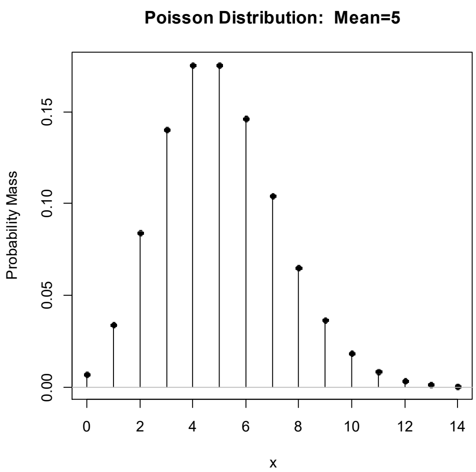
Graphs of Poisson Distribution

- λ=0.5

λ=1







Joint Probability Function

- A joint probability function is used to express the probability that X takes the specific value x and simultaneously Y takes the value y, as a function of x and y.
- $P(x,y)=P(X=x \cap Y=y)$

Practice

- Coin toss – 3 times
- X: # of heads on the 1st toss
- Y: total number of heads
- S={hhh, hht, hth, htt, thh, tht, tth, ttt}
- Complete the table below, joint probabilities and marginal probabilities.

		Y (total number of heads in 3 tosses)				P(X)
		0	1	2	3	
X (# of heads on 1st toss)	0					
	1					
P(Y)						

		Y (total number of heads in 3 tosses)				P(X)
		0	1	2	3	
X (# of heads on 1st toss)	0	{ttt} =1/8	{tht, tth} =2/8	{thh} =1/8	0	=4/8
	1	0	{htt} =1/8	{hht, hth} =2/8	{hhh} =1/8	=4/8
P(Y)		=1/8	=3/8	=3/8	=1/8	

S={hhh, hht, hth, htt, thh, tht, tth, ttt}

Conditional Probability Function

- The conditional probability function of the random variable Y expresses the probability that Y takes the value y when the value x is specified for X.

$$P(y|x) = \frac{P(x,y)}{P(x)} \qquad P(x|y) = \frac{P(x,y)}{P(y)}$$

		Y (total number of heads in 3 tosses)				P(X)
		0	1	2	3	
X (# of heads on 1st toss)	0	{ttt} =1/8	{tht, tth} =2/8	{thh} =1/8	0	P(X=0)= 4/8
	1	0	{htt} =1/8	{hht, hth} =2/8	{hhh} =1/8	P(X=1)= 4/8
P(Y)		P(Y=0) =1/8	P(Y=1) =3/8	P(Y=2) =3/8	P(Y=3)= 1/8	

- Find P(X=1|Y=2) and P(Y=2|X=1)

$$P(x|y) = \frac{P(x,y)}{P(y)} \qquad P(y|x) = \frac{P(x,y)}{P(x)}$$

Independence

- The jointly distributed random variables X and Y are said to be independent iff their joint probability function is the product of their marginal probability functions.

- P(x,y) = P(x)P(y) for all possible pairs of values x and y.

$\Rightarrow$P(y|x) = [P(x)P(y)]/P(x) = P(y)

$\Rightarrow$P(x|y)=[P(x)P(y)]/P(y) = P(x)

		Y (total number of heads in 3 tosses)				P(X)
		0	1	2	3	
X (# of heads on 1st toss)	0	{ttt} =1/8	{tht, tth} =2/8	{thh} =1/8	0	=4/8
	1	0	{htt} =1/8	{hht, hth} =2/8	{hhh} =1/8	=4/8
P(Y)		=1/8	=3/8	=3/8	=1/8	

- Check if X and Y are independent.
- (Independence: P(x,y) = P(x)P(y) for all possible pairs of values x and y.)
- P(x=0,y=0)=?=P(x=0)P(y=0)
- 1/8 =?= (4/8)*(1/8) = 4/64=1/16   Not independent.
- (If you find one pair which is not satisfying the equality, you can state that they are not independent.)

Mean and Variance

- E(X)=X1*P(X1)+X2*P(X2)…+Xn*P(Xn)
- Var(X)=(X1-E(X))^2*P(X1)+(X2-E(X))^2*P(X2)+…(Xn-E(X))^2*P(Xn)
- Or Var(X)= X1^2*P(X1)+X2^2*P(X2)+…Xn^2*P(Xn)-E(X)^2

Covariance

- Let X and Y be discrete random variables with means μx and μy.

$$\mathrm{Cov}(X, Y) = E\left[(x - \mu_x)\left(y - \mu_y\right)\right] = \sum_x \sum_y (x - \mu_x)(y - \mu_y)P(x, y)$$

$$\mathrm{Cov}(x, y) = E[XY] - \mu_x \mu_y = \sum_x \sum_y xyP(x, y) - \mu_x \mu_y$$

Correlation

$$\rho = \mathrm{Cor}(x, y) = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}$$

- $-1 \leq \rho \leq 1$
- $\rho = 0$: no linear relationship between x and y.
- $\rho > 0$: positive linear relationship between x and y
- $\rho = 1$: perfectly positive linear relationship
- $\rho < 0$: negative linear relationship between x and y
- $\rho = -1$: perfectly negative linear relationship

Linear Function of Several Random Variables

- W = aX+bY
- E(W)= $a\mu_X + b\mu_Y$

$$Var(W) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X,Y)$$
$$Var(W) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCorr(X,Y)\sigma_X\sigma_Y$$

- W=aX-bY
- E(W)= $a\mu_X - b\mu_Y$

$$Var(W) = a^2\sigma_X^2 + b^2\sigma_Y^2 - 2abCov(X,Y)$$
$$Var(W) = a^2\sigma_X^2 + b^2\sigma_Y^2 - 2abCorr(X,Y)\sigma_X\sigma_Y$$

Practice

- Setting:
- New model of a cell phone
- Advertised on a TV program.

- 15% of people who watch the show regularly and could identify the product.
- 16% of people watch the show regularly
- 45% of people could identify the product.

- X=1: regularly watch the show, X=0 otherwise
- Y=1: correctly identify the product, Y=0 otherwise

A table for joint probability functions

		Y(Identify Product)		
		0	1	P(X)
X(watch the show regularly)	0			
	1			
	P(Y)			1

- 15% of people who watch the show regularly and could identify the product.
- 16% of people watch the show regularly
- 45% of people could identify the product.

	0	1	P(X)
0	0.54	0.30	0.84
1	0.01	0.15	0.16
P(Y)	0.55	0.45	1

- e.g. $P(0,0)=P(X=0 \cap Y=0)=0.54$

Practice

a. Find the joint probability function of X and Y.
b. Find conditional probability function of Y given X=1.
c. Are X and Y independent?
d. Find Cov(X,Y) and Cor(X,Y)
e. If W = 2X + Y, Find E(W) and Var(W)
f. If W = 2X – Y, Find E(W) and Var(W)

	0	1	P(X)
0	0.54	0.30	0.84
1	0.01	0.15	0.16
P(Y)	0.55	0.45	1

b. Find conditional probability function of Y given X=1.

$P(Y=0|X=1)=P(X=1 \cap Y=0)/P(X=1) = 0.01/0.16=0.0625$

$P(Y=1|X=1)=P(X=1 \cap Y=1)/P(X=1)=0.15/0.16 =0.9375$

c. Are X and Y independent?

- If independent, the following should be true:
- P(X,Y)=P(X|Y)P(Y)=P(X)P(Y)   for all combinations of X and Y.
- For example, P(0,0)=P(X=0|Y=0)P(Y=0)=P(X=0)P(Y=0) ???
- P(0,0)=0.54 from the table.
- P(X=0)P(Y=0)=0.84*0.55=0.462. Since they are not equal, X and Y are NOT independent.

c. Find COV(X,Y) and CORR(X,Y)

- E(X)=0*0.84+1*0.16=0.16
- E(Y)=0*0.55+1*0.45=0.45
- Var(X)=(0-0.16)^2*0.84+(1-0.16)^2*0.16=0.1344
- Var(Y)=(0-0.45)^2*0.55+(1-0.45)^2*0.45=0.2475
- COV(X,Y)=0*0*0.54+0*1*0.30+1*0*0.01+1*1*0.15-0.16*0.45=0.078
- Cor(X,Y)=0.078/sqrt(0.1344)*sqrt(0.2475)=0.4277

$$\rho = \mathrm{Cor}(x,y) = \frac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_y} \qquad \mathrm{Cov}(x,y) = E[XY] - \mu_x \mu_y = \sum_x \sum_y xyP(x,y) - \mu_x \mu_y$$

e. If W = 2X + Y, Find E(W) and Var(W)

- E(W)=2*E(X)+E(Y)
=2*0.16+1*0.45=0.77

- Var(W)=(2^2)*Var(X)+(1^2)*Var(Y)+2*2*1*Cov(X,Y)
=4*0.1344+1*0.2475+4*0.078=1.0971

$$Var(W) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X,Y)$$

f. If W = 2X - Y, Find E(W) and Var(W)

- E(W)=2*0.16-1*0.45=-0.13
- Var(W)=4*0.1344+1*0.2475-4*0.078=0.4731

$$Var(W) = a^2\sigma_X^2 + b^2\sigma_Y^2 - 2abCov(X,Y)$$

Continuous Probability Distribution

Continuous Distributions covered.

- Uniform
- Normal
- Exponential

Continuous Random Variable

- A variable that can assume any value in an interval

e.g.
Time required to complete a task
Temperature of a solution
Thickness, height, weight

Probability Density Function (PDF): f(x)

- f(x) > 0 for all values of x.
- Area under f(x) over all values of x = 1.
- P(a<x<b) = area under f(x) between a and b

$$= \int_a^b f(x)dx$$

Cumulative Distribution Function (CDF): F(x)

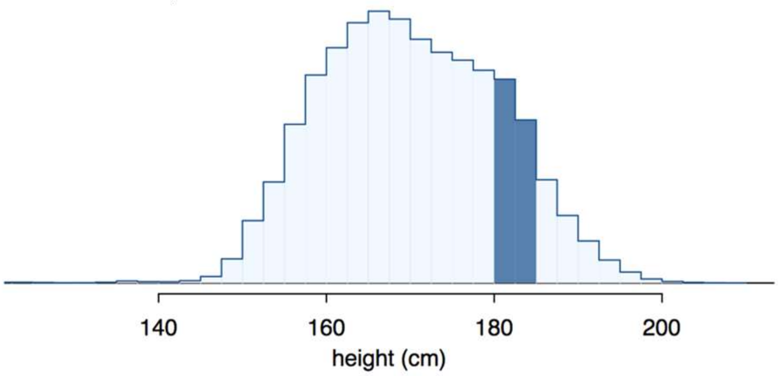
$$F(x) = P(X \leq x)$$

$$P(a < X < b) = F(b) - F(a)$$

$$F(x_0) = \int_{x_{min}}^{x_0} f(x)dx$$

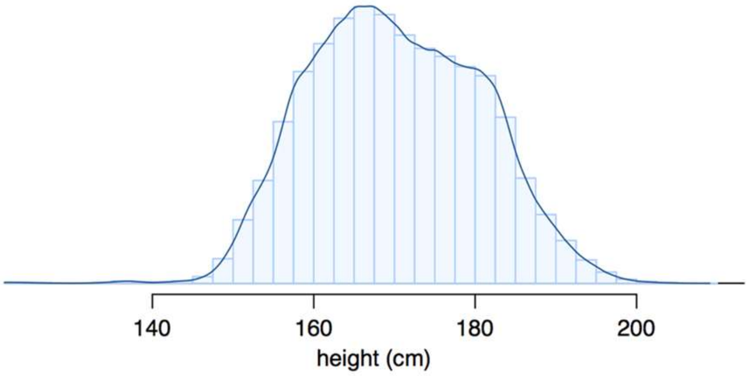
Continuous distributions

- Below is a histogram of the distribution of heights of US adults.
- The proportion of data that falls in the shaded bins gives the probability that a randomly sampled US adult is between 180 cm and 185 cm (about 5'11" to 6'1").



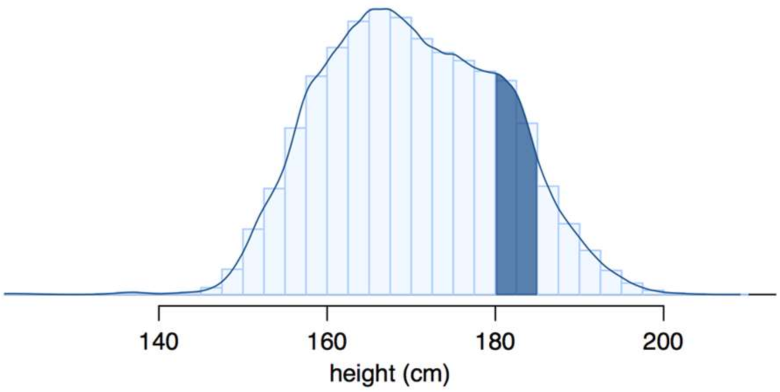
From histograms to continuous distributions

Since height is a continuous numerical variable, its probability density function is a smooth curve.



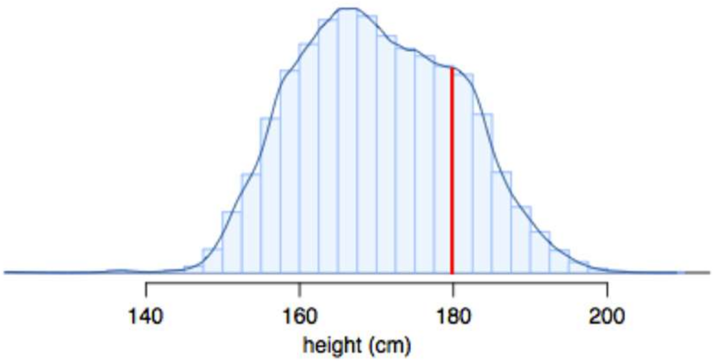
Probabilities from continuous distributions

Therefore, the probability that a randomly sampled US adult is between 180 cm and 185 cm can also be estimated as the shaded area under the curve.



By definition...

Since continuous probabilities are estimated as "the area under the curve", the probability of a person being exactly 180 cm (or any exact value) is defined as 0.



Mean, Variance, Standard Deviation of Continuous R.V.

$$\mu_x = E(X) = \int_x xf(x)dx$$

$$\sigma_x^2 = E[(X - \mu_x)^2]$$

$$= \int_x (X - \mu_x)^2 f(x)dx$$

$$\sigma_x^2 = E(X^2) - \mu_x^2$$

$$= \int_x X^2 f(x)dx - \mu_x^2$$

Linear Function of Variables

- Let W = a + bX, then

$$\mu_W = E(a+bX) = a + bE(X)$$
$$= a + b\mu_x$$

$$\sigma_W^2 = Var(a+bX) = b^2\sigma_x^2$$

$$\sigma_W = |b|\sigma_x$$

Practice

- Y=290-5T
- E(T)=24, stdev(T)=4

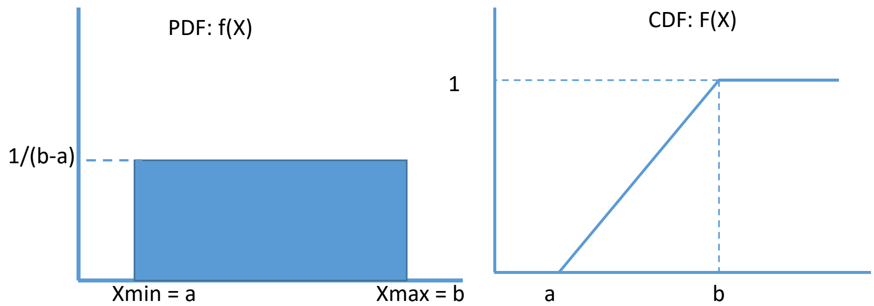
Q: Find E(Y) and stdev(Y)

E(Y)=290-5*24=290-120=170.
stdev(Y)=5*4=20

Uniform Distribution

- X~ U(a,b)

Has equal probabilities for all possible outcomes of the random variable.



$$PDF: f(x) = \begin{cases} \dfrac{1}{b-a} \; for \; a \le X \le b \\ 0 \; for \; X < a, X > b \end{cases}$$

$$CDF: F(x) = \begin{cases} 0 \; for \; X < a \\ \dfrac{x-a}{b-a} \; for \; a \le X \le b \\ 1 \; for \; x > b \end{cases}$$

X ~ U(a, b)

<= R.V. has Uniform Distribution on interval [a b]

- Find PDF if X ~ U(2,6)

=> f(x) = 1/(6-2) = 0.25 for $2 \leq X \leq 6$

$$PDF: f(x) = \begin{Bmatrix} \frac{1}{b-a} \; for \; a \leq X \leq b \\ 0 \; for \; X < a, X > b \end{Bmatrix}$$

Practice

$$PDF: f(x) = \begin{Bmatrix} \frac{1}{b-a} \; for \; a \leq X \leq b \\ 0 \; for \; X < a, X > b \end{Bmatrix}$$

$$CDF: F(x) = \begin{Bmatrix} 0 \; for \; X < a \\ \frac{x-a}{b-a} \; for \; a \leq X \leq b \\ 1 \; for \; x > b \end{Bmatrix}$$

- Oil pipeline over 2km
- Probability of finding any fracture is equal over 2 km => X~ U(0,2).

a. Find PDF for X.
b. Draw PDF.
c. Find CDF for X.
d. Draw CDF.
e. Find the probability that any fracture is found between 0.5 and 1.5 km.

a. Find PDF for X. => f(x)=1/2 for [0 2]
b. Draw PDF.
c. Find CDF for X. => F(x)=x/2 for [0 2]
d. Draw CDF.
e. Find the probability that any fracture is found between 0.5 and 1.5 km.
=> F(1.5)-F(0.5)=1.5/2 – 0.5/2 = 0.5

Practice

$$PDF: f(x) = \begin{Bmatrix} \dfrac{1}{b-a} \; for\; a \le X \le b \\ 0\; for\; X < a, X > b \end{Bmatrix}$$

$$CDF: F(x) = \begin{Bmatrix} 0\; for\; X < a \\ \dfrac{x-a}{b-a} \; for\; a \le X \le b \\ 1\; for\; x > b \end{Bmatrix}$$

- For X~ U(0, 4),

a. Find PDF
b. Draw PDF
c. Find and Draw CDF.
d. Find the probability that X being between 0 and 1.
e. Find the probability that X being between 0 and 0.5, and 3.5 and 4.

a. Find PDF => f(x)=1/4 =0.25 for 0<x<4.
b. Draw PDF
c. Find and Draw CDF. =>F(X)= (1/4)*X=0.25X.
d. Find the probability that X being between 0 and 1. => F(1)-F(0)=0.25-0.
e. Find the probability that X being between 0 and 0.5 and 3.5 and 4.
=> F(0.5)-F(0)=0.25*0.5 – 0.25*0 = 0.125
F(4)-F(3.5)=0.25*4 – 0.25*3.5=0.125
=> = 0.25

Mean and Variance of X ~ U(a,b)

$$\mu_x = \frac{a+b}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$