CENG 551 Probability and Stochastic Processes for Engineers

Week # 8

Random Variables: Poisson RV's

Dr. Deniz Ozdemir
Deniz.ozdemir@yasar.edu.tr
Web: dozdemir.yasar.edu.tr

Plan for the Session

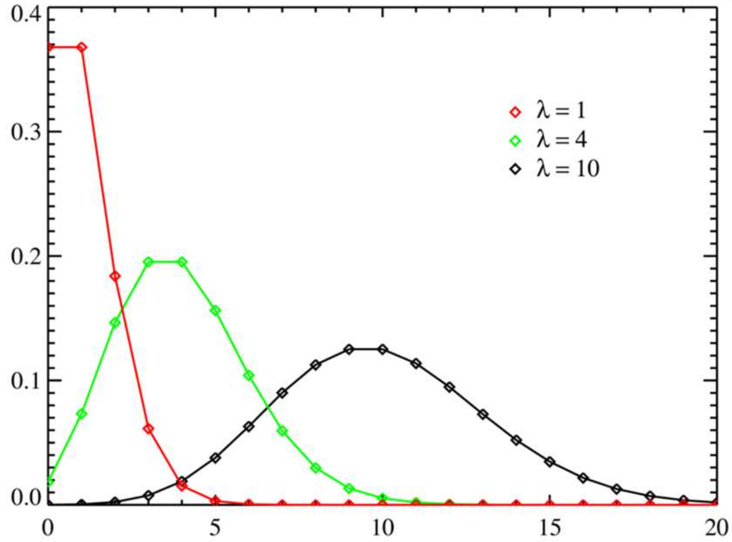
- Poisson Random Variables
- Continuous Random Variables
  - Multiple random variables
  - Conditioning
  - Independence

Poisson Distribution

- The probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event.
  - Number of misprints in a page/book
  - Number of customers do shopping from 7/11 on a given day
  - Number of tacos sold in a particular taqueria each day
  - Number of wrong numbers that are dialed in a day
- If the expected number of occurrences in this interval is λ, then the probability that there are exactly k occurrences (k = 0, 1, 2, ...) is equal to

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

PMF



Properties of Poisson Distribution (λ)

- PMF: $P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$
- Expectation:
  - $E[X] =$
  - $E[X^2] =$
- Variance:
  - $E(X^2) - (E[X])^2 =$

Law of Small Numbers (rare events)

- The events being counted are actually the outcomes of discrete trials, and would more precisely be modelled using the *binomial distribution*.
- Binomial(n, λ/n) $\rightarrow$Poisson (λ) as $n \rightarrow \infty$.
- This provides a means by which to approximate random variables using the Poisson distribution rather than the more-cumbersome binomial distribution.
- This limit is sometimes known as the **law of rare events**, since each of the individual Bernoulli events rarely triggers.
  - the total count of success events in a Poisson process need not be rare if the parameter λ is not small.
  - i.e., the number of telephone calls to a busy switchboard in one hour follows a Poisson distribution with the events appearing frequent to the operator, but they are rare from the point of the average member of the population who is very unlikely to make a call to that switchboard in that hour.

Example: Poisson Distribution

- Suppose that the probability that an item produced by a certain machine will be defective is 0.1. What is the probability that a sample of 10 items will contain at most 1 defective item?

  P(0 defected item) + P (1 defected item)

  Binomial Solution: $\binom{10}{0}.1^0.9^{10} + \binom{10}{0}.1^1.9^9 = 0.7361$

  Poisson Approximation: $e^{-1}\frac{\lambda^0}{0!} + e^{-1}\frac{\lambda^1}{1!} \approx 0.7358$

Birthday Problem

- At a gathering of s randomly chosen students what is the probability that at least 2 will have the same birthday?

P(at least 2 have same birthday)=1-P(all s students have different birthdays).

Assume 365 days in a year. Think of students' birthdays as a sample of these 365 days.

The total number of possible outcomes is:

N=365^s (ordered, with replacement)

The number of ways that s students can have different birthdays is

M=364!/(365- s)! (ordered, without replacement)

P(all s students have different birthdays) =

M / N .

For s = 20, P(all s students have different birthdays) = 0.58856

Birthday Problem : Poisson Approximation

- At a gathering of n randomly chosen students what is the probability that at least 2 will have the same birthday?

Suppose we have a trial of all pairs of individuals i and j (i≠j)

Trial (i, j) is success if i and j have same birthday

E_{ij}: Event that i and j have same birthday, $P(E_{ij})$ = 1/365

Total number of trials $\binom{n}{2} = \frac{n(n-1)}{2}$

Approximate $\lambda = np = n(n\text{-}1)/730$

P(all n students have different birthdays)

= P(0 success) = $P(X=0) = e^{-n(n-1)/730}\frac{\lambda^0}{0!}$

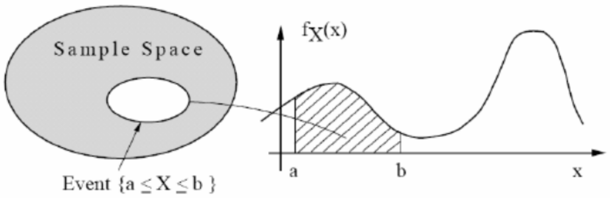
For n = 20, P(0 success) = $e^{-380/730} \approx 0.5942$

For n = 20, P(all s students have different birthdays) = 0.58856

11/13/2012 9

Continuous Random Variables

Continuous Random Variables (PDF)



$$\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)dx$$

$$\boxed{\mathbf{P}(x \le X \le x+\delta) \approx f_X(x) \cdot \delta}$$

Probability density function

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Means and Variance

- Analogous to discrete version:

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx$$

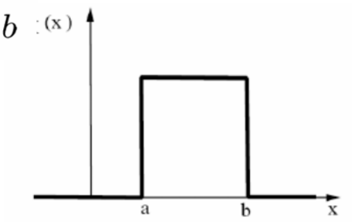
$$\mathbf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$$

$$\text{var}(X) = \sigma_X^2$$

$$= \int_{-\infty}^{\infty} (x - \mathbf{E}[X])^2 \cdot f_X(x)dx$$

Example: Uniform distribution

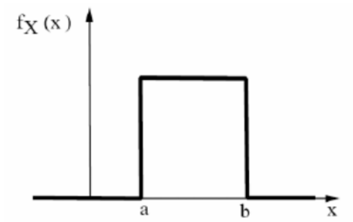
- PDF: $f_X(x) = \frac{1}{b-a} \qquad a \le x \le b$
- Expectation:
  - $E[X] = \frac{a+b}{2}$
- Variance:
  - $E(X - E[X])^2 =$

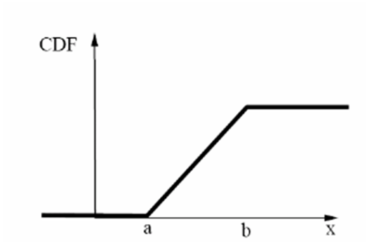


$$\int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}$$

Cumulative Distribution Function

- CDF: $F_X(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f_X(t)dt$
- <u>Uniform Distribution</u>



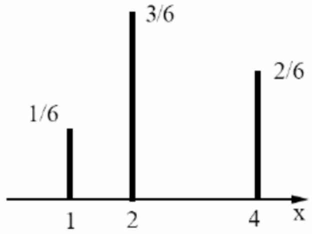


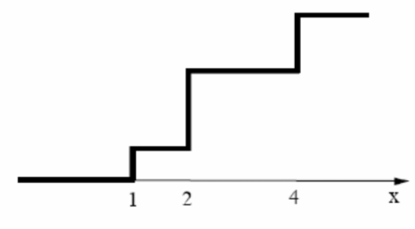
Cumulative Distribution Function

- CDF for discrete functions:

$$F_X(x) = \mathbf{P}(X \le x) = \sum_{k \le x} p_X(k)$$

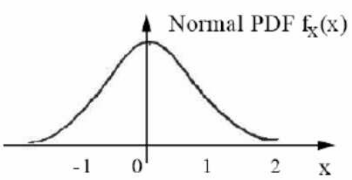
- <u>Example:</u>





Standard Gaussian (Normal) PDF

- Standard Normal: $N(0, 1)$
- PDF: $$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$
- Expectation: $E[X] = 0$
- Variance: $\mathrm{Var}(X)=1$



General Gaussian (Normal) PDF

General Normal: $N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

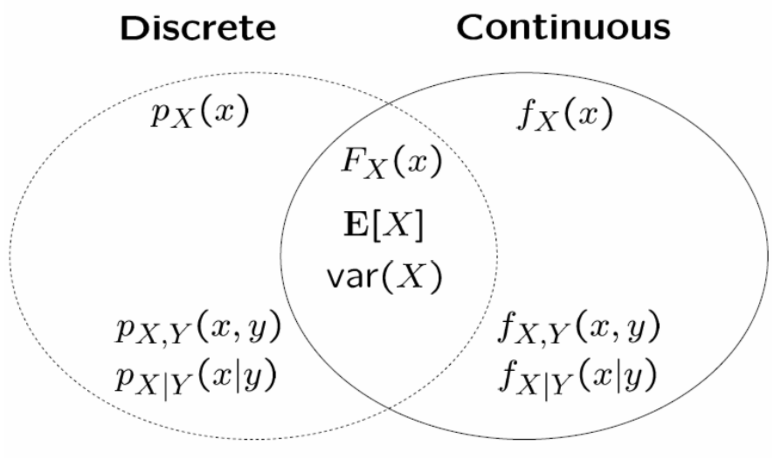
Calculating Normal Probabilities

- No closed form available for CDF.
  - But, there are tables (for standard normal).
- If $X \sim N(\mu, \sigma^2)$ then $N(0, 1)$
- So if $X \sim N(2, 16)$ then $P(X \leq 3) =$

  $P(X \leq 3) = P((X-2)/4 \leq (3-2)/4)$

  $= P(Z \leq 0.25) =$

Summary of concepts



Joint PMF

$$p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y)$$

y \ x	1	2	3	4
4	1/20	2/20	2/20	
3	2/20	4/20	1/20	2/20
2		1/20	3/20	1/20
1		1/20		

$$\boxed{p_X(x) = \textstyle\sum_y p_{X,Y}(x,y)}$$

$$p_{X|Y}(x|y)$$
$$= \mathbf{P}(X = x | Y = y)$$
$$= \frac{p_{X,Y}(x,y)}{p_Y(x)}$$

$$\textstyle\sum_x \sum_y p_{X,Y}(x,y) = 1$$

$$\textstyle\sum_x p_{X|Y}(x|y) = 1$$

Joint PDF $f_{X,Y}(x, y)$

$$p_{X,Y}(x,y) = \mathbf{P}(X = x \text{ and } Y = y)$$

$f_{X,Y}(x,y)$

$$\mathbf{P}(A) = \iint_A f_{X,Y}(x,y)dxdy$$

- Interpretation

$$\mathbf{P}(x \le X \le x+\delta, y \le Y \le y+\delta) \approx f_{X,Y}(x,y) \cdot \delta^2$$

- Expectation

$$\mathbf{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy$$

Joint PDF $f_{X,Y}(x, y)$ II

$$\mathbf{P}(A) = \iint_A f_{X,Y}(x,y)dxdy$$

- From the joint to the marginal:

$$f_X(x) \cdot \delta \approx \mathbf{P}(x \le X \le x+\delta) =$$

$$\boxed{\int_{-\infty}^{\infty}\int_{x}^{x+\delta} f_{X,Y}(t,y)dtdy \approx \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \cdot \delta}$$

- X and Y are called independent iff:

$$\boxed{f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)}$$

Joint PMF: Independence

$$p_{X,Y,Z}(x,y,z) = p_X(x)p_{Y|X}(y|x)p_{Z|X,Y}(z|x,y)$$

- Random variables X, Y and Z are independent if (for all x, y and z):

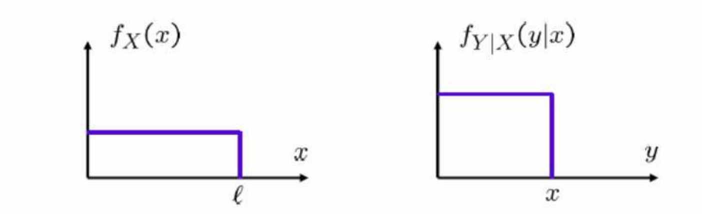
$$\boxed{p_{X,Y,Z}(x,y,z) = p_X(x) \cdot p_Y(y) \cdot p_Z(z)}$$

Conditioning

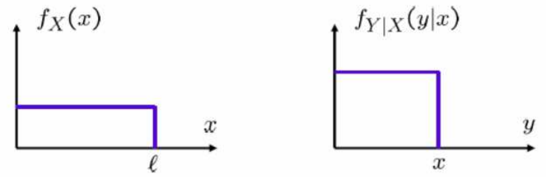
- Recall, again: $\mathbf{P}(x \le X \le x + \delta) \approx f_X(x) \cdot \delta$
- By analogy: $\mathbf{P}(x \le X \le x + \delta | Y \approx y) \approx f_{X|Y}(x|y) \cdot \delta$
- Thus, the definition: $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$
- Conditioning is a “section”of the joint PDF, normalized.
- Independence gives: $f_{X|Y}(x|y) = f_X(x)$

Example: Stick-Breaking

- Break a stick of length ℓ twice:
  - X: first break point, chosen uniformly between 0 and ℓ.
  - Y: second break point, chosen (given $X=x$) uniformly from 0 to x.

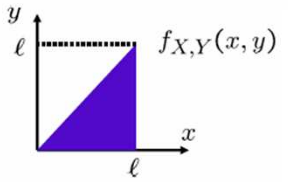


Example: Stick-Breaking: *Joint PDF*

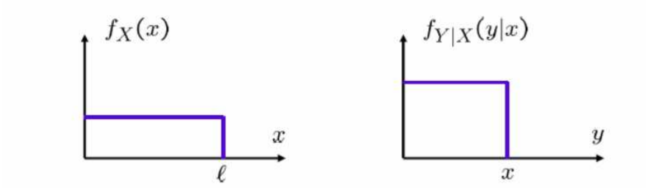


- Joint PDF: $f_{X,Y}(x,y) = f_X(x) \cdot f_{Y|X}(y|x)$

$$= \frac{1}{\ell x} \qquad 0 \leq y < x \leq \ell$$



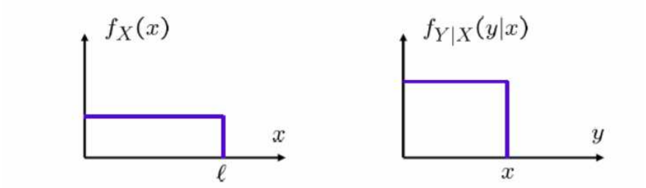
Example: Stick-Breaking: *Marginal PDF*



- M-PDF: $f_Y(y) = \int f_{X,Y}(x,y)\,dx$

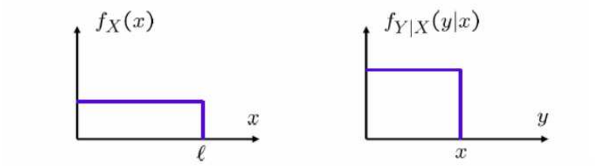
$$= \int_y^{\ell} \frac{1}{\ell x}\,dx = \frac{1}{\ell}\log\frac{\ell}{y}, \qquad 0 \le y \le \ell$$

Example: Stick-Breaking: *Expectation*



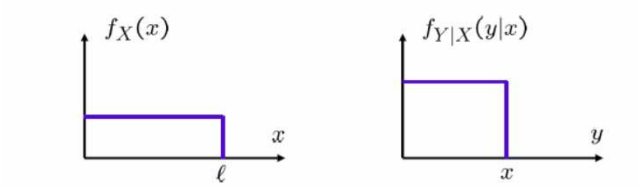
- Conditional Expectation of Y, given $X=x$:
  - $\mathbf{E}[Y|X=x] = \int y f_{Y|X}(y \mid X=x)\,dy = \frac{x}{2}$
- Expectation of Y:
  - $\mathbf{E}[Y] =$

Example: Stick-Breaking: *Expectation*

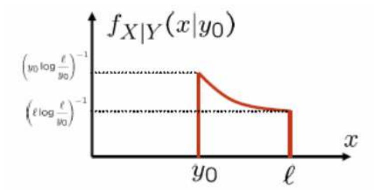


$$\mathbf{E}[Y] = \int_0^{\ell} y f_Y(y)\, dy \; = \; \int_0^{\ell} y \frac{1}{\ell} \log \frac{\ell}{y}\, dy = \frac{\ell}{4}$$

Example: Stick-Breaking: *Conditional PDF*

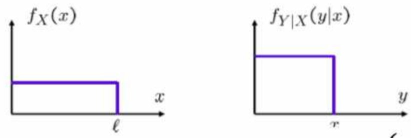


- $f_{X|Y}(x \,|Y{=}y) =$

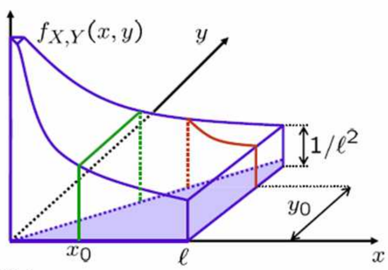


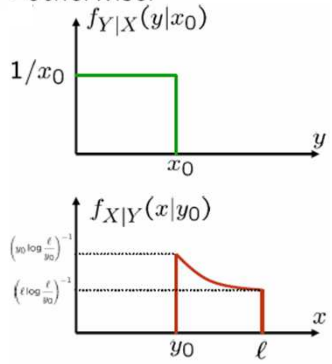
Example: Stick-Breaking:

Conditioning "slices" the joint PDF



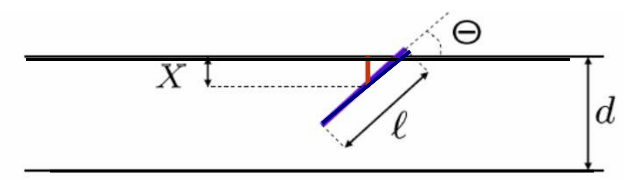
- Recall the joint PDF: $f_{X,Y}(x,y) = \begin{cases} \frac{1}{\ell x} & 0 \le y < x \le \ell \\ 0 & \text{otherwise.} \end{cases}$
- Pictorially:





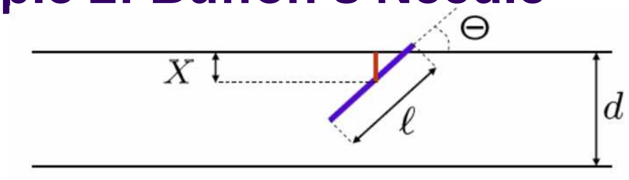
Example 2: Buffon's Needle

- Parallel lines at distance d
- Needle of length ℓ (assume $\ell < d$)
- Find **P**(needle intersects one of the lines).



- Midpoint-nearest line distance: $X \in [0. d/2]$
- Needle-lines acute angle: $\Theta \in [0,\pi/2]$

Example 2: Buffon's Needle



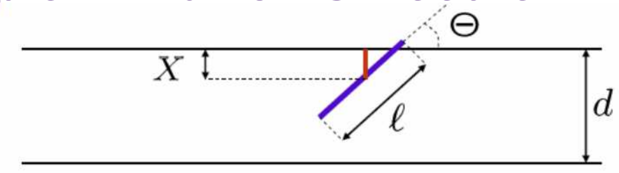
- Model: X, Θ uniform and independent

$$f_{X,\Theta}(x,\theta) = f_X(x)\cdot f_\Theta(\theta)$$
$$= \frac{2}{d}\cdot\frac{2}{\pi} \quad 0\le x\le d/2,\ 0\le\theta\le\pi/2$$

- When does the needle intersect a line?

If $X \le \frac{\ell}{2}\sin\Theta$

Example 2: Buffon's Needle



$$\mathbf{P}\left(X\le\frac{\ell}{2}\sin\Theta\right) = \int\int_{x\le\frac{\ell}{2}\sin\theta} f_X(x)f_\Theta(\theta)\,dx\,d\theta$$
$$= \frac{4}{\pi d}\int_0^{\pi/2}\int_0^{(\ell/2)\sin\theta} dx\,d\theta$$
$$= \frac{4}{\pi d}\int_0^{\pi/2}\frac{\ell}{2}\sin\theta\,d\theta = \frac{2\ell}{\pi d}$$

Next Time..

- MIDTERM

On Nov. 27 - Tuesday

- More on continuous r.v.s
- Derived distributions