國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 16:
review
section 4: further topics on random variable
李祈均

Review on Probabilistic Inference

Variations in Bayes rule

- X is unknown, but we have some prior belief in how X is distributed
- We observe random variable Y
- Need a model of the box:
  - If the true state of the world is X, how do we expect Y to be distributed
  - Inference problem, knowing Y, what can we say about X?
  - **Inference problem is about finding probability distribution**
    - $P(X|Y)$ (PMF or PDF)

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}$$

$$P_Y(y) = \sum_x p_X(x) p_{Y|X}(y|x)$$

- Say, this is the discrete case
- Typical example:
  - $X = 1,0$: airplane present/no present
  - $Y = 1,0$: something did/did not register on radar
- Inference problem:
  - $P(X|Y)$: given the radar measurement, calculate the probability that the plane is up there
  - **Make decision this way (find X=x that maximizes P(X|Y))**

- Continuous case:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

$$f_Y(y) = \int_x f_X(x)\, f_{Y|X}(y|x) dx$$

- Typical example
  - X: some signal, e.g., current through resistor, with $f_X(x)$
  - Y: signal with some noise, e.g., Gaussian noise
- Inference problem is the same
  - Compute $f_{X|Y}$
    - **Make decision this way (find X=x that maximizes some criterion of $f_{X|Y}$, maybe we can find E(X|Y)?)**
  - $f_{Y|X}$ is the model of signal with noise

Discrete X, Continuous Y

- Typical example (communication)
  – X: discrete signal (say a bit, 0,1),
  – Y: measured signal (X + Gaussian noise)

- prior: $P_X(x)$
  - Could be equally-likely to send 0,1 (PMF)

– Continuous noise model: $f_{Y|X}(y|x)$
  - Conditional densities of y in a universe that's specified by a particular value of x
  - This is something that you know

$$P(X = x|Y = y) \approx P(X = x|y \leq Y \leq y + \delta)$$

$$= \frac{P(X = x)P(y \leq Y \leq y + \delta|X = x)}{P(y \leq Y \leq y + \delta)}$$

$$\approx \frac{P(X = x)f_{Y|X}(y|x)\delta}{f_Y(y)\delta}$$

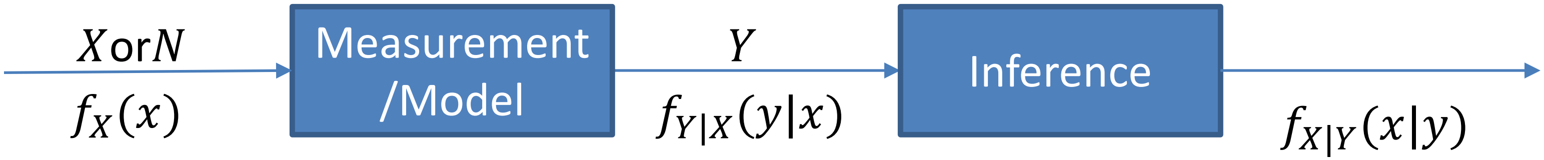
$$= \frac{P_X(x)f_{Y|X}(y|x)}{f_Y(y)}$$

$$f_Y(y) = \sum_i P_X(i)f_{Y|X}(y|i)$$

Put it together:

$$P_{X|Y}(x|y) = \frac{p_X(x)f_{Y|X}(y|x)}{f_Y(y)}$$

Look at it again: Bayes rule



- If Y is a continuous random variable, we have:
  - If X is continuous random variable

$$f_{X,Y}(x,y) = f_Y(y)f_{X|Y}(x|y) = f_X(x)f_{Y|X}(y|x)$$

So,

$$f_{X|Y}(x|y) = \frac{f_X(x)f_{Y|X}(y|x)}{f_Y(y)} = \frac{f_X(x)f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(t)f_{Y|X}(y|t)dt}$$

- If N is discrete random variable, we have

$$f_Y(y)P(N=n|Y=y) = p_N(n)f_{Y|N}(y|n)$$

Resulting in the formula:

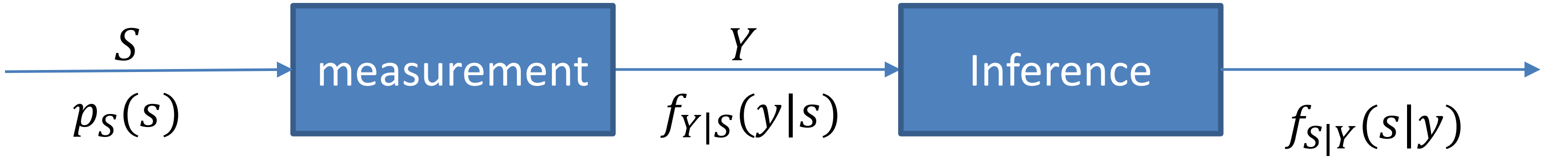
$$P(N=n|Y=y) = \frac{p_N(n)f_{Y|N}(y|n)}{f_Y(y)} = \frac{p_N(n)f_{Y|N}(y|n)}{\sum_i p_N(i)f_{Y|N}(y|i)}$$

A quick example

- A Binary signal S is transmitted, and we are given that

$$P(S = 1) = p, P(S = -1) = 1 - p$$

The received signal is $Y = N + S$, where N is Gaussian noise (μ=0, σ^2=1), independent of S. What is the probability that S=1, as a function of the observed value y of Y?



$$P(S = 1|Y = y) = \frac{p_S(1) f_{Y|S}(y|1)}{f_Y(y)}$$

$$f_Y(y) = \sum_i p_N(i) f_{Y|N}(y|i)$$

What is $f_{Y|S}$?

If the measurement is additive Gaussian noise ($N(0,1)$) to the original signal *S*, that is $Y = S + Noise$

$$f_{Y|S}(y|S = 1) \sim N\left(\mu = 1, \sigma^2 = 1\right)$$

$$f_{Y|S}(y|S = -1) \sim N(\mu = -1, \sigma^2 = 1)$$

$$f_{Y|S}(y|S=1) \sim N\left(\mu = 1, \sigma^2 = 1\right) = \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2}$$

$$f_{S|Y}(S=1|Y=y) = \frac{p_S(1) f_{Y|S}(y|1)}{f_Y(y)}$$

$$\frac{p * \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2}}{\frac{p}{\sqrt{2\pi}} e^{-(y-1)^2/2} + \frac{1-p}{\sqrt{2\pi}} e^{-(y+1)^2/2}}$$

$$= \frac{pe^{y}}{pe^{y} + (1-p)e^{-y}}$$

With this probability density function, we can do ‘informed’ probabilistic inference, e.g., finding the probability that your received signal y is within a particular range

Example 1

- A light bulb produced by the company is known to have an exponentially distributed lifetime Y. However, the company has been experiencing quality control problems. On any given day, the parameter λ of the PDF of Y is actually a random variable, uniformly distributed in the interval $[1,3/2]$.

- We test a light bulb and record its lifetime. What can we say about the underlying parameter λ

- Let’s model the parameter, λ, as a uniform random variable Λ with PDF:

$$f_{\Lambda}(\lambda) = 2, \text{for } 1 \leq \lambda \leq \frac{3}{2}$$

So if we test the light bulb and record it lifetime $(Y = y)$, then how can we use that ($Y = y$) to modify our original belief of $f_{\Lambda}(\lambda)$

Therefore, the available information about Λ is captured by the conditional PDF $f_{\Lambda|Y}(\lambda|Y = y)$, we can compute it using continuous Bayes rule

$$f_{\Lambda|Y}(y) = \frac{f_{\Lambda}(\lambda) f_{Y|\Lambda}(y|\lambda)}{\int_{-\infty}^{\infty} f_{\Lambda}(t) f_{Y|\Lambda}(y|t) dt} = \frac{2\lambda e^{-\lambda y}}{\int_{1}^{3/2} 2te^{-ty} dt}, \text{for } 1 \leq \lambda \leq \frac{3}{2}$$

This is statistics – somewhat, we will come back to this

Think about it, after you measure it, what’s the most likely parameter for this exponential distribution

Getting parameter’s PDF (function) estimated using real world data, then you pick a number (scalar, realization) that you think is right!

Continuous random variable

- Function (PDF, CDF, so on)
  - Know about exponential distribution
  - Know how to compute Gaussian distribution's probability
- Expected value, variances
- Joint PDF
- Marginal PDF
- Conditional PDF
  - Each can be operated with expectation operator
- Bayesian inference

Outline

Reading 4.1

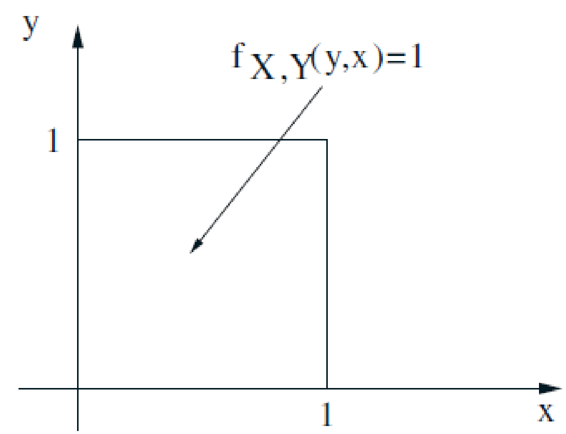
- Derived distributions

Derived distribution

Idea:

- Trying to find the distribution of a function of one or more known random variable with known probability law
- These types of distributions are called *derived distribution*

- Say for example:
  - You are told (X,Y) is uniform
  - You are interested in finding X/Y
    - Finding the ratio



Derived distribution

- Essentially, you are trying to obtain the pdf for:

$$g(X,Y) = \frac{Y}{X}$$

  - Involves deriving a distribution from known distribution
  - Note that $g(X,Y)$ is also a random variable

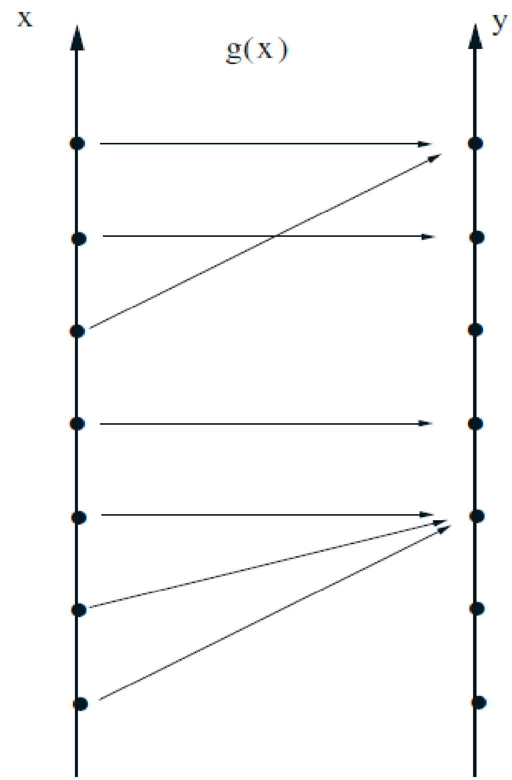
- *Important: When not to find them
  - If you don't need to find $g(X,Y)$ if not needed
  - For example:

$$E[g(X,Y)] = \iint g(x,y) f_{X,Y}(x,y)\,dx\,dy$$

How to find them?

- Discrete case
  - Think about random variable is a mapping function
  - Obtain probability mass for each possible value of $Y = g(X)$

$$p_Y(y) = P(g(X) = y) = \sum_{x:g(x)=y} p_X(x)$$



- Just sum up the probability in the X that leads to $Y = y$ to come up with tl new PMF

Continuous case

- Can we do the same thing as in the discrete case?
  - Not really, we don’t have probability that operates “points-to-points”
  - Because?
  - Probability in continuous case is zero at every individual point
- So what we want to find is not the probability,
  - We want to find the actual density for Y
  - In essence:
    - Find the probability of falling in a small interval around $Y = y$, and look back at the set of x’s that leads to those y’s
    - Then set those probability equal to each other to find density

Continuous case

- Instead of working with intervals, we work directly with CDF
  - What is CDFs? Cumulative distribution function
  - Essentially, a big intervals!

- Two step procedure
  - Get CDF of Y: $F_Y(y) = P(Y \leq y)$
  - Differentiate the CDF to obtain PDF

$$f_Y(y) = \frac{dF_Y}{dy}(y)$$

Example

- X: uniform on [0,2]
- $Y = X^3$
  - Now how to find the PDF for Y?

- Y can take value from [0,8]
  - You might think, since all x's are equally-likely, y's should be equally-likely too?
  - Is that really correct?

Example

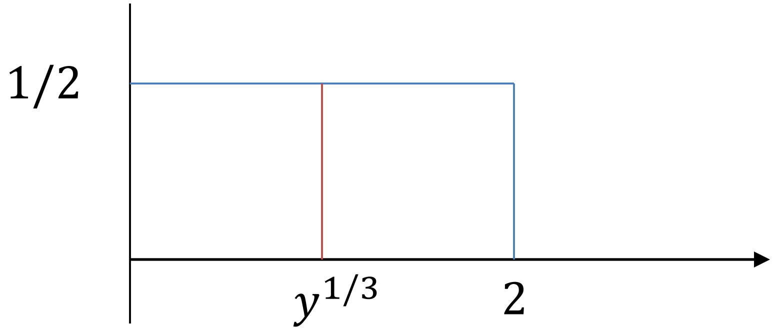
- Lets try to apply the two-step procedure

$$F_Y(y) = P(Y \leq y) = P\left(X^3 \leq y\right)$$

$$= P\left(X \leq y^{1/3}\right)$$

What is this, X is uniform, and this is cumulative distribution in the interval of $\left[0, y^{1/3}\right]$

- $P\left(X \le y^{1/3}\right)$
  - What is the probability (essentially the area)



  - Area: $\frac{1}{2} y^{1/3}$

$$F_Y(y) = P\left(X \le y^{1/3}\right) = \frac{1}{2} y^{1/3}$$

$$f_Y(y) = \frac{dF_Y}{dy}(y) = \frac{1}{6y^{\frac{2}{3}}}, 0 \le y \le 8$$

This is not a uniform distribution!

Another example

- You are driving from Boston to New York. You are on cruise control, hence your speed is uniformly distributed between 30 and 60 mph.

- You are going to drive a distance of 200 miles, and the time to take for the trip is 200/V
  - $T(V) = \frac{200}{V}$
  - $V \sim UNIFORM(30,60)$
  - Now we have distribution of V, can we find the distribution of T?

- $P(T \leq t) = P\left(\frac{200}{V} \leq t\right) = P(\frac{200}{t} \leq V)$

$$f_V(v) = \begin{cases} \frac{1}{30}, 30 \leq v \leq 60 \\ 0, \text{otherwise} \end{cases}$$

The CDF of $F_V(v) = P(V \leq v)$

$$F_V(v) = \begin{cases} 0, \text{if v} \leq 30 \\ \frac{v-30}{30}, \text{if } 30 \leq v \leq 60 \\ 1, \text{if } 60 \leq v \end{cases}$$

$$P\left(\frac{200}{t} \leq V\right) = 1 - F_V\left(\frac{200}{t}\right)$$

$$= \begin{cases} 0, \text{if } t \leq 200/60 \\ 1 - \dfrac{\frac{200}{t} - 30}{30}, \text{if } 200/60 \leq t \leq 200/30 \\ 1, \text{if } 200/30 \leq t \end{cases}$$

- Now that we have the CDF, need to find PDF
  - Just perform differentiation

$$F_T(t) = \begin{cases} 0, \text{if } t \leq \dfrac{200}{60} \\ 1 - \dfrac{\frac{200}{t} - 30}{30}, \text{if } \dfrac{200}{60} \leq t \leq \dfrac{200}{30} \\ 1, \text{if } \dfrac{200}{30} \leq t \end{cases}$$

$$f_T(t) = \begin{cases} 0, \text{if } t \leq \dfrac{200}{60} \\ \dfrac{20}{3t^2}, \text{if } \dfrac{200}{60} \leq t \leq \dfrac{200}{30} \\ 0, \text{if } \dfrac{200}{30} \leq t \end{cases}$$

Some note

- The first problem:
  - Monotonically increasing function
    - Find CDF directly

- The second problem:
  - Monotonically decreasing function
    - The inequality has to be reverted at some point (note the 1-CDF)

Linear case

- Lets say Y is linear function of X

$$Y = 2X + 5$$

- Imagine simple intuition to reason:
  - Stretch and rescale

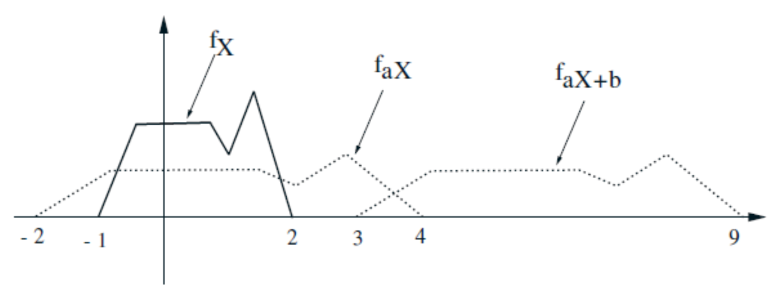
From figure above,

Let's write the formula first (in the case of Y is linear function of X: $Y = aX + b$)

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

- Multiplying means stretching

- Shifting just moving the pdfs



Let's derive this using two step approach

Assume $a > 0$

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y)$$

$$= P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

Now, let's differentiate this,

Using chain rule:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

- With $a < 0$

$$F_Y(y) = P(Y \le y) = P(aX + b \le y)$$

$$= P\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)$$

Now, let's differentiate this,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$