Probability review

Alejandro Ribeiro
Dept. of Electrical and Systems Engineering
University of Pennsylvania
aribeiro@seas.upenn.edu
http://www.seas.upenn.edu/users/~aribeiro/

September 11, 2015

Conditional probabilities

Conditional pmf and cdf for discrete RVs

- Recall definition of conditional probability for events E and F

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

- Change in likelihoods when information is given, renormalization

- Define the conditional pmf of RV X given Y as (both RVs discrete)

$$p_{X|Y}(x \mid y) = \mathrm{P}\left[X = x \mid Y = y\right] = \frac{\mathrm{P}\left[X = x, Y = y\right]}{\mathrm{P}\left[Y = y\right]}$$

- Can rewrite as $\Rightarrow p_{X|Y}(x \mid y) = \dfrac{\mathrm{P}\left[X = x, Y = y\right]}{\mathrm{P}\left[Y = y\right]} = \dfrac{p_{XY}(x, y)}{p_Y(y)}$
- Pmf for random variable x, given parameter y ("Y not random anymore")

- Define conditional cdf as (a range of X conditional on a value of Y)

$$F_{X|Y}(x \mid y) = \mathrm{P}\left[X \leq x \mid Y = y\right] = \sum_{z \leq x} p_{X|Y}(z \mid y)$$

Example

- Independent Bernoulli Y and Z, variable $X = Y + Z$
- Conditional pmf of X given Y? For $X = 0,\ Y = 0$

$$p_{X|Y}(X = 0 \,\big|\, Y = 0) = \frac{\mathrm{P}[X = 0, Y = 0]}{\mathrm{P}[Y = 0]} = \frac{(1-p)^2}{1-p} = 1 - p$$

- Or, from joint and marginal pdfs (just a matter of definition)

$$p_{X|Y}(X = 0 \,\big|\, Y = 0) = \frac{p_{XY}(0,0)}{p_Y(0)} = \frac{(1-p)^2}{1-p} = 1 - p$$

- Can compute the rest analogously

$$\begin{aligned} &p_{X|Y}(0|0) = (1-p), && p_{X|Y}(1|0) = p, && p_{X|Y}(2|0) = 0 \\ &p_{X|Y}(0|1) = 0, && p_{X|Y}(1|1) = 1 - p, && p_{X|Y}(2|1) = p \end{aligned}$$

Conditional pdf and cdf for continuous RVs

- Define conditional pdf of RV X given Y as (both RVs continuous)

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

- For motivation, define intervals $\Delta x = [x, x+dx]$ and $\Delta y = [y, y+dy]$
- Can approximate conditional probability $\mathrm{P}\left[X \in \Delta x \,\middle|\, Y \in \Delta y\right]$ as

$$\mathrm{P}\left[X \in \Delta x \,\middle|\, Y \in \Delta y\right] = \frac{\mathrm{P}\left[X \in \Delta x, Y \in \Delta y\right]}{\mathrm{P}\left[Y \in \Delta y\right]} \approx \frac{f_{XY}(x, y)dxdy}{f_Y(y)dy}$$

- From definition of conditional pdf it follows after simplifying terms

$$\mathrm{P}\left[X \in \Delta x \,\middle|\, Y \in \Delta y\right] \approx f_{X|Y}(x \,|\, y)dx$$

- Which is what we would expect of a density
- Conditional cdf defined as $\Rightarrow F_{X|Y}(x) = \int_{-\infty}^{x} f_{X|Y}(u \,|\, y)du$

Example: Communications channel

- Random message (RV) Y, transmit signal y (realization of Y)
- Received signal is $x = y + z$ (z realization of random noise)
- Can model communication system as a relation between RVs

$$X = Y + Z$$

- Model communication noise as $Z \sim \mathcal{N}(0, \sigma^2)$ independent of Y

- Conditional pdf of X given Y. Use definition:

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

- Problem is we don't know $f_{XY}(x, y)$. Have to calculate

- Computing conditional probs. typically easier than computing joints

- If $Y = y$ is given, then "Y not random anymore" (Dorothy's principle)

  $\Rightarrow$ It still is random in reality, we are thinking of it as given
- If Y were not random, say $Y = y$ with y given then ...

$$X = y + Z$$

- Cdf of X, now easily obtained

$$\mathrm{P}\left[X \leq x\right] = \mathrm{P}\left[y + Z \leq x\right] = \mathrm{P}\left[Z \leq x - y\right] = \int_{-\infty}^{x-y} p_Z(z)\, dz$$

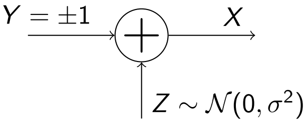
- But since Z is normal with 0 mean and variance σ^2

$$\mathrm{P}\left[X \leq x\right] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x-y} e^{-z^2/2\sigma^2}\, dz = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(z-y)^2/2\sigma^2}\, dz$$

- X is normal with mean y and variance σ^2

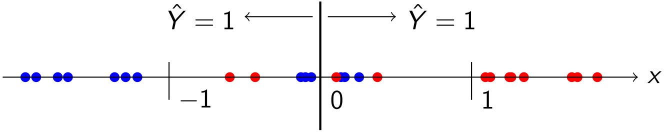
Digital communications channel

- Conditioning is a common tool to compute probabilities



- Message 1 (prob. p) $\Rightarrow$ Transmit $Y = 1$
- Message 2 (prob. q) $\Rightarrow$ Transmit $Y = -1$
- Received signal $\Rightarrow X = Y + Z$

- Decoding rule $\Rightarrow \hat{Y} = 1$ if $X \geq 0$, $\hat{Y} = -1$ if $X < 0$
- What is the probability of error, $P_e := \mathrm{P}\left[\hat{Y} \neq Y\right]$?
- Red dots to the left and blue dots to the right are errors

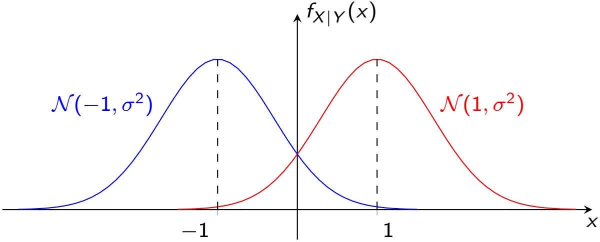


- From communications channel example we know
- If $Y = 1$, then $X \sim \mathcal{N}(1, \sigma^2)$, conditional pdf is

$$f_{X|Y}(x, 1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-1)^2/2\sigma^2}$$

- If $Y = -1$, then $X \sim \mathcal{N}(-1, \sigma^2)$, conditional pdf is

$$f_{X|Y}(x, -1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x+1)^2/2\sigma^2}$$



Error probability

- Write probability of error by conditioning on $Y = \pm 1$ (total probability)

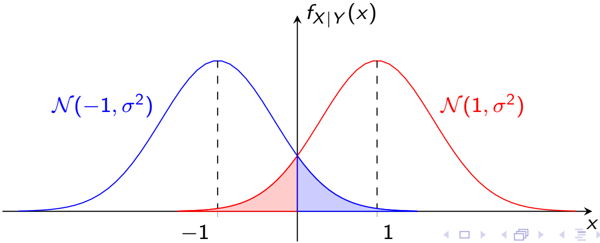
$$P_e = \mathrm{P}\{\hat{Y} \neq Y \mid Y = 1\}\mathrm{P}\{Y = 1\} + \mathrm{P}\{\hat{Y} \neq Y \mid Y = -1\}\mathrm{P}\{Y = -1\}$$
$$= \mathrm{P}\{\hat{Y} = -1 \mid Y = 1\}\ p \qquad + \mathrm{P}\{\hat{Y} = 1 \mid Y = -1\}\ q$$

- But according to the decision rule

$$P_e = \mathrm{P}\{X < 0 \mid Y = 1\}p + \mathrm{P}\{X \geq 0 \mid Y = -1\}q$$

- But X given Y is normally distributed, then

$$P_e = \frac{p}{\sqrt{2\pi}\sigma}\int_0^\infty e^{-(x-1)^2/2\sigma^2} + \frac{q}{\sqrt{2\pi}\sigma}\int_{-\infty}^0 e^{-(x+1)^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{-1} e^{-x^2/2\sigma^2}$$



Conditional probabilities

Conditional expectation

Definition of conditional expectation

- For continuous RVs X, Y define conditional expectation as

$$\mathbb{E}\left[X \mid y\right] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

- For discrete RVs X, Y conditional expectation is

$$\mathbb{E}\left[X \mid y\right] = \sum_{x} x\, p_{X|Y}(x|y)$$

- Defined for given y $\Rightarrow \mathbb{E}\left[X \mid y\right]$ is a value
- All possible values y of Y $\Rightarrow$ random variable $\mathbb{E}\left[X \mid Y\right]$
- Y is RV, $\mathbb{E}\left[X \mid y\right]$ value associated with outcome $Y = y$

Double expectation

- If $\mathbb{E}\left[X \,\middle|\, Y\right]$ is a RV, can compute expected value $\mathbb{E}_Y\left[\mathbb{E}_X\left(X \,\middle|\, Y\right)\right]$
  - Subindices are for clarity purposes, innermost expectation is with respect to X, outermost with respect to Y
- What is $\mathbb{E}_Y\left[\mathbb{E}_X\left(X \,\middle|\, Y\right)\right]$? Not surprisingly $\Rightarrow \mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}_X\left(X \,\middle|\, Y\right)\right]$
- Show for discrete RVs (write integrals for continuous)

$$\mathbb{E}_Y\left[\mathbb{E}_X\left(X \,\middle|\, Y\right)\right] = \sum_y \mathbb{E}_X\left(X \,\middle|\, y\right) p_Y(y) = \sum_y \left[\sum_x x\, p_{X|Y}(x|y)\right] p_Y(y)$$

$$= \sum_x x \left[\sum_y p_{X|Y}(x|y) p_Y(y)\right] = \sum_x x \left[\sum_y p_{X,Y}(x,y)\right]$$

$$= \sum_x x p_X(x) = \mathbb{E}[X]$$

- Yields a method to compute expected values

$\Rightarrow$ Condition on $Y = y$ $\Rightarrow X \,\middle|\, y$

$\Rightarrow$ Compute expected value over X for given y $\Rightarrow \mathbb{E}_X\left(X \,\middle|\, y\right)$

$\Rightarrow$ Compute expected value over all values y of Y $\Rightarrow \mathbb{E}_Y\left[\mathbb{E}_X\left(X \,\middle|\, Y\right)\right]$

Example

- Seniors get $A = 4$ with prob. 0.5, $B = 3$ with prob. 0.5
- Juniors get $B = 3$ with prob. 0.6, $B = 2$ with prob. 0.4
- Exchange student's standing: senior (junior) with prob. 0.7 (0.3)
- Expectation of X = exchange student's grade?

- Begin conditioning on standing

$$\mathbb{E}\left[X \,\middle|\, \text{Senior}\right] = 0.5 \times 4 + 0.5 \times 3 = 3.5$$
$$\mathbb{E}\left[X \,\middle|\, \text{Junior}\right] = 0.6 \times 3 + 0.4 \times 2 = 2.6$$

- Now sum over standing's probability

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[X \,\middle|\, \text{Senior}\right] \mathrm{P}[\text{Senior}] + \mathbb{E}\left[X \,\middle|\, \text{Junior}\right] \mathrm{P}[\text{Junior}] \\ &= 3.5 \times 0.7 + 2.6 \times 0.3 \\ &= 3.23\end{aligned}$$

Conditioning to compute expectations

- As with probabilities conditioning is useful to compute expectations.

  $\Rightarrow$ Spreads difficulty into simpler problems

Example

- A baseball player hits X_i runs per game
- Expected number of runs is $\mathbb{E}[X_i] = \mathbb{E}[X]$ independently of game
- Player plays N games in the season. N is random (playoffs, injuries?)
- Expected value of number of games is $\mathbb{E}[N]$
- What is the expected number of runs in the season ? $\Rightarrow \mathbb{E}\left[\sum_{i=1}^{N} X_i\right]$
- Both, N and X_i are random

Sum of random number of random quantities

Step 1: Condition on $N = n$ then

$$\left[\left(\sum_{i=1}^{N} X_i\right) \Big| \, N = n\right] = \sum_{i=1}^{n} X_i$$

Step 2: Compute expected value with respect to X_i

$$\mathbb{E}_{X_i}\left[\left(\sum_{i=1}^{N} X_i\right) \Big| \, N = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}\left[X\right]$$

Second equality possible because n is a number (not a RV like N)

Step 3: Conpute expected value with respect to values n of N

$$\mathbb{E}_N\left[\mathbb{E}_{X_i}\left[\left(\sum_{i=1}^{N} X_i\right) \Big| \, N\right]\right] = \mathbb{E}_N\left[N\mathbb{E}\left[X\right]\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[X\right]$$

Yielding result $\Rightarrow \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] = \mathbb{E}\left[N\right]\mathbb{E}\left[X\right]$