Probability Theory
Lecture 2: Conditioning

Hector Rodriguez-Deniz

STIMA, IDA
Linköping University

Overview Lecture 2

- **Conditional distributions**
- **Conditional expectation, conditional variance**
- **Distributions with random parameters, Bayesian approach**
- **Regression and Prediction**

Conditional distributions

- A and B events, $P(B) > 0$

  Conditional probability of A given B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- A and B are **independent** if and only if $P(A|B) = P(A)$
- X and Y jointly distributed **discrete** random variables, $p_X(x) > 0$

  Conditional pmf of Y given $X = x$

$$p_{Y|X=x}(y) = p(Y = y|X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

Conditional distributions

- X and Y jointly distributed **continuous** random variables, $f_X(x) > 0$
  Conditional pdf of Y given $X = x$:

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

- **Conditional cdf** of Y given $X = x$:

$$F_{Y|X=x}(y) = \begin{cases} \sum_{z \leq y} p_{Y|X=x}(z) & \text{if } Y \text{is discrete} \\ \int_{-\infty}^{y} f_{Y|X=x}(z)dz & \text{if } Y \text{is continuous} \end{cases}$$

- Example: $X \sim U(0,1)$, $Y|X = x \sim U(0,x)$

Conditional expectation

- **Conditional expectation** of Y given $X = x$:

$$E(Y|X=x) = \begin{cases} \sum_y y \cdot p_{Y|X=x}(y) & \text{if } Y \text{is discrete} \\ \int_{-\infty}^{\infty} y \cdot f_{Y|X=x}(y)dy & \text{if } Y \text{is continuous} \end{cases}$$

- $h(x) = E(Y|X=x)$ - function of x

  $h(X) = E(Y|X)$ - random variable

  $h(X) = E(Y|X)$ depends only on X

- X and Y independent

$$E(Y|X) = E(Y)$$

Law of iterated expectation

- Let $E(|Y|) < \infty$
- **Law of iterated expectation**:

$$E\left[E(Y|X)\right] = E(Y)$$

- Expectation $E(Y|X)$ is with respect to $f_{Y|X}(y)$

  Expectation $E\left[E(Y|X)\right]$ is with respect to $f_X(x)$
- Proof for discrete case - Ex. 2 in Seminar 2

Conditional variance

数形结合，在确定x的情况下来思考y的分布

- **Conditional variance** of Y given $X = x$:

$$Var(Y|X = x) = E\left[(Y - E(Y|X = x))^2 \,|X = x\right]$$

- $v(X) = Var(Y|X)$ - random variable
- $v(X) = Var(Y|X)$ depends on X
- If $E(Y^2) < \infty$

$$Var(Y) = E\left[Var(Y|X)\right] + Var\left[E(Y|X)\right]$$

Distributions with random parameters

- X - random variable with distribution depending on unknown parameter θ
- Consider θ to be random variable
- Example 1:
  - $X|N = n \sim Bin(n, p)$ and $N \sim Po(\lambda)$
  - Number of trials is random: $N \sim Po(\lambda)$
  - Marginal distribution of X - ?

这个是把可能选择的n的数量进行了意义列举

- For discrete random variables use Law of total probability:

$$P(A) = \sum_{j=1}^{n} P(A|H_j)P(H_j)$$

where $H_i \cap H_j = \emptyset$, $H_1 \cap \cdots \cap H_n = \Omega$, $A \in \Omega$

Distributions with random parameters

- Example 2:
  - $X|Y = \lambda \sim N(0, 1/\lambda)$ and $Y \sim \Gamma\left(\frac{n}{2}, \frac{2}{n}\right)$
  - Variance parameter is random: $Y \sim \Gamma\left(\frac{n}{2}, \frac{2}{n}\right)$
  - Marginal distribution for X - ?
- For continuous random variables use formula

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|\theta=a}(x) f_\theta(a)\, da$$

Formula follows from $f_{X|\theta=a}(x) f_\theta(a) = f_{X,\theta}(x, \theta)$

- Pdf of $Y \sim \Gamma(p, a)$:

$$f_Y(y) = \frac{1}{\Gamma(p)} y^{p-1} \frac{1}{a^p} e^{-y/a},\ x > 0$$

where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ - Gamma function, $a, p > 0$

Bayesian coin tossing

- X_n - number of heads after n trials

$$X_n | P = p \sim Bin(n, p)$$

- **Prior distribution**: $P \sim U(0, 1)$
- **Posterior distribution**:

$$P | X_n = k \sim Beta(k + 1, n + 1 - k)$$

- Marginal distribution of X_n:

$$X_n \sim U(\{0, 1, 2, ..., n\})$$

Regression and Prediction

- $X_1, \ldots, X_n$ and Y - jointly distributed random variables
- **Regression function**:

$$h(\mathrm{x}) = h(x_1, ..., x_n) = E(Y|X_1 = x_1, ..., X_n = x_n) = E(Y|\mathrm{X} = \mathrm{x})$$

- **Predictor**: $\hat{Y} = d(\mathrm{X})$
- **Linear predictor**: $d(\mathrm{X}) = a_0 + a_1 X_1 + ... a_n X_n$
- **Mean squared error** (MSE): $E\left[Y - d(\mathrm{X})\right]^2$
- The **best predictor** of Y [minimizes MSE]: $E(Y|\mathrm{X} = \mathrm{x})$
- Best **linear** predictor in case $n = 1$:

$$\boxed{\hat{Y}} = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(X - \mu_x)$$

- (X, Y) jointly normal $\Rightarrow$ $\boxed{E(Y|X = x) \text{ linear}}$

Thank you for your attention!