Lecture Stat 302
Introduction to Probability - Slides 21

AD

April 2010

Conditional Distributions: Discrete Case

- Given two r.v. X, Y, we have

Discrete	Continuous
$p(x, y)$	$f(x, y)$
$p_{X\|Y}(x\| y) = \frac{p(x,y)}{p_Y(y)}$	$f_{X\|Y}(x\| y) = \frac{f(x,y)}{f_Y(y)}$
$E(g(X)\| y) = \sum g(x).p_{X\|Y}(x\| y)$	$E(g(X)\| y) = \int g(x).f_{X\|Y}(x\| y)\, dx$

Example: A Toy example

- Let X and Y be two r.v. of joint p.d.f.

$$f(x,y) = \begin{cases} 1/x & \text{for } 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

- (a) Establish the expression of $f_X(x)$.
- (b) Establish the expression of $f_Y(y)$.
- (c) Establish the expression of $f_{X|Y}(x|y)$ and $E(X|Y=y)$.
- (d) Establish the expression of $f_{Y|X}(y|x)$ and $E(Y|X=x)$.

Example: A Toy example

- (a) We have for $0 \leq x \leq 1$

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \frac{1}{x} \int_0^x dy = 1$$

and $f_X(x) = 0$ elsewhere.

- (b) We have for $0 \leq y \leq 1$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_y^1 \frac{1}{x} dx = -\log y$$

and $f_Y(y) = 0$ elsewhere.

Example: A Toy example

- (c) We have for $0 \leq y \leq x \leq 1$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = -\frac{1}{x \log y}$$

and $f_{X|Y}(x|y) = 0$ elsewhere. Hence

$$E(X|Y = y) = \int_y^1 x \; f_{X|Y}(x|y)\, dx = -\frac{(1-y)}{\log y}.$$

- (d) We have for $0 \leq y \leq x \leq 1$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{x}.$$

and $f_{Y|X}(y|x) = 0$ elsewhere. Hence

$$E(Y|X = x) = \int_0^x y \; f_{Y|X}(y|x)\, dy = \frac{x}{2}.$$

Example: Bayesian Signal Estimation

- Let X be a random signal. We do not observe X directly but have access to a noisy measurement

$$Y = X + N$$

  where N is a random noise. Assume that X is a normal r.v. of parameters $\left(m, \sigma_x^2\right)$ and N is a normal of mean 0 and variance σ_n^2, X and N are independent. Show that the conditional pdf $f_{X|Y}\left(x|\, y\right)$ is normal. Compute $E\left(X|y\right)$ and $Var\left(X|y\right)$.

- We have $f_X\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\left(x-m\right)^2/\left(2\sigma_x^2\right)}$ and $f_{Y|X}\left(y|\, x\right) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\left(y-x\right)^2/\left(2\sigma_n^2\right)}$ so

$$\begin{aligned} f_{X|Y}\left(x|\, y\right) &= \frac{f_{Y|X}\left(y|\, x\right) f_X\left(x\right)}{f_Y\left(y\right)} \\ &= \frac{\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\left(x-m\right)^2/\left(2\sigma_x^2\right)} \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\left(y-x\right)^2/\left(2\sigma_n^2\right)}}{f_Y\left(y\right)} \end{aligned}$$

Example: Signal Estimation

- We have

$$\frac{(x-m)^2}{\sigma_x^2}+\frac{(y-x)^2}{\sigma_n^2}=\left(\frac{1}{\sigma_x^2}+\frac{1}{\sigma_n^2}\right)x^2-2\left(\frac{y}{\sigma_n^2}+\frac{m}{\sigma_x^2}\right)x+\frac{y^2}{\sigma_n^2}+\frac{m^2}{\sigma_x^2}$$
$$=\frac{1}{\sigma^2}(x-\mu)^2-\frac{\mu^2}{\sigma^2}+\frac{y^2}{\sigma_n^2}+\frac{m^2}{\sigma_x^2}$$

  where

$$\sigma^2 = \frac{\sigma_x^2\sigma_n^2}{\sigma_x^2+\sigma_n^2},$$
$$\mu = \sigma^2\left(\frac{y}{\sigma_n^2}+\frac{m}{\sigma_x^2}\right)=\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}y+\frac{\sigma_n^2}{\sigma_x^2+\sigma_n^2}m$$

- Hence it follows that

$$f_{X|Y}(x|y)\propto e^{-(x-\mu)^2}/\left(2\sigma^2\right)$$

  and $f_{X|Y}(x|y)$ can only be a normal density of mean $E(X|y)=\mu$ and $Var(X|y)=\sigma^2$.

Example: Bayesian Signal Estimation

- As $\sigma_n^2 \to 0$, the observation $Y = y$ is very informative about X and we have
$$\sigma^2 = \frac{\sigma_x^2 \sigma_n^2}{\sigma_x^2 + \sigma_n^2} \approx \frac{\sigma_x^2 \sigma_n^2}{\sigma_x^2} = \sigma_n^2$$
and
$$\mu = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2} y + \frac{\sigma_n^2}{\sigma_x^2 + \sigma_n^2} m \approx y.$$
- As $\sigma_n^2 \to \infty$, the observation $Y = y$ is not informative about X and we have
$$\sigma^2 = \frac{\sigma_x^2 \sigma_n^2}{\sigma_x^2 + \sigma_n^2} \approx \frac{\sigma_x^2 \sigma_n^2}{\sigma_n^2} = \sigma_x^2$$
and
$$\mu = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2} y + \frac{\sigma_n^2}{\sigma_x^2 + \sigma_n^2} m \approx m.$$

Example: Sequential Bayesian Signal Estimation

- Assume now that you have access at each discrete time index k to

$$Y_k = X + N_k$$

where $\{N_k\}_{k\geq 1}$ are independent normal of mean 0 and variance σ_n^2.

- At time k, we can compute similarly $f_{X|(Y_1,Y_2,\ldots,Y_k)}(x|\, y_1, y_2, \ldots, y_k)$

$$\begin{aligned} & f_{X|(Y_1,Y_2,\ldots,Y_k)}(x|\, y_1, y_2, \ldots, y_k) \\ = \quad & \frac{f_{(Y_1,Y_2,\ldots,Y_k)|X}(y_1, y_2, \ldots, y_k|\, x)\, f_X(x)}{f_{(Y_1,Y_2,\ldots,Y_k)}(y_1, y_2, \ldots, y_k)} \\ \propto \quad & e^{-(x-m)^2/\left(2\sigma_x^2\right)} e^{-\sum_{j=1}^{k}(y-x_k)^2/\left(2\sigma_n^2\right)} \end{aligned}$$

which is a normal of mean $E(X|y_1, y_2, \ldots, y_k) = \mu_k$ and $Var(X|y_1, y_2, \ldots, y_k) = \sigma_k^2$.

Example: Sequential Bayesian Signal Estimation

- We can also compute $f_{X|(Y_1,Y_2,\ldots,Y_k)}(x|\, y_1, y_2, \ldots, y_k)$ recursively

$$
\begin{aligned}
& f_{X|(Y_1,Y_2,\ldots,Y_k)}(x|\, y_1, y_2, \ldots, y_k) \\
= \quad & \frac{f_{Y_k|X}(y_k|\, x)\, f_{X|(Y_1,Y_2,\ldots,Y_{k-1})}(x|\, y_1, y_2, \ldots, y_{k-1})}{f_{Y_k|(Y_1,Y_2,\ldots,Y_{k-1})}(y_k|\, y_1, y_2, \ldots, y_{k-1})}
\end{aligned}
$$

- Hence using calculations similar to the simple case, we obtain by substituting $m \leftarrow \mu_{k-1}$ and $\sigma_x^2 \leftarrow \sigma_{k-1}^2$

$$
\sigma_k^2 = \frac{\sigma_{k-1}^2 \sigma_n^2}{\sigma_{k-1}^2 + \sigma_n^2}, \;\; \mu_k = \sigma_k^2 \left(\frac{y_k}{\sigma_n^2} + \frac{\mu_{k-1}}{\sigma_{k-1}^2} \right)
$$

  with $\mu_0 = m$, $\sigma_0^2 = \sigma_x^2$.
- This is a special case of the Kalman(-Stratonovich) filter: one of the most popular algorithms in applied mathematics/aerospace/control/telecommunications/econometrics etc.

- We have
$$E\left[E\left(X|Y\right)\right]=E\left[X\right].$$
- **Proof**. We have
$$\begin{aligned}
& E\left[E\left(X|Y\right)\right] \\
= & \int\left[\int x.f_{X|Y}\left(x|y\right)dx\right]f_Y\left(y\right)dy \\
= & \int\int x.\underbrace{f_{X|Y}\left(x|y\right)f_Y\left(y\right)}_{=f\left(x,y\right)}dxdy \\
= & \int\int x.f_X\left(x\right)f_{Y|X}\left(y|x\right)dxdy=\int\left[\int x.f_X\left(x\right)dx\right]f_{Y|X}\left(y|x\right)dy \\
= & \int x.f_X\left(x\right)dx=E\left[X\right]
\end{aligned}$$
- This is valid for any function $E\left[E\left(g\left(X\right)|Y\right)\right]=E\left[g\left(X\right)\right]$.

Example: Workplace Injuries

- The number of workplace injuries N occuring in a factory on any given day is Poisson distributed of parameter λ. The parameter λ is a random variable determined by the level of activity in the factory and is uniformly distributed on the interval $[0, 3]$. What is the expectation of N?
- We have
$$E[N] = E[E[N|\lambda]]$$
where
$$E[N|\lambda] = \lambda.$$
- Hence we have
$$E[N] = E[\lambda] = \frac{3}{2}.$$

Properties of the Conditional Variance

- We have the following variance decomposition

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

hence as $E[Var(X|Y)] \geq 0$

$$Var[E(X|Y)] \leq Var[X]$$

- **Proof**. We have

$$\begin{aligned} Var[E(X|Y)] &= E\left(\{E(X|Y)\}^2\right) - E(E(X|Y))^2 \\ &= E\left(\{E(X|Y)\}^2\right) - E(X)^2 \end{aligned}$$

and

$$Var(X|Y) = E\left(X^2\middle| Y\right) - \{E(X|Y)\}^2$$

so

$$\begin{aligned} E[Var(X|Y)] &= E\left(E\left(X^2\middle| Y\right)\right) - E\left(\{E(X|Y)\}^2\right) \\ &= E\left(X^2\right) - E\left(\{E(X|Y)\}^2\right) \end{aligned}$$

Example: Workplace Injuries

- The number of workplace injuries N occuring in a factory on any given day is Poisson distributed of parameter λ. The parameter λ is a random variable determined by the level of activity in the factory and is uniformly distributed on the interval $[0, 3]$. What is the variance of N?
- We have

$$Var[N] = Var[E(N|\lambda)] + E[Var(N|\lambda)]$$

where

$$E(N|\lambda) = Var(N|\lambda) = \lambda$$

as, given λ, N is a Poisson random variable of param λ.
- Hence

$$\begin{aligned} Var[N] &= Var[\lambda] + E[\lambda] \\ &= \frac{3^2}{12} + \frac{3}{2} = \frac{9}{4} \end{aligned}$$

as λ follows an uniform distribution on $[0, 3]$.

Optimality of Conditional Expectation

- Let us consider two r.v. X and Y. Assume we observe Y and want to find a way to estimate X based on Y. Then in some sense, $E(X|Y)$ is the best possible estimate of X.
- **Proposition.** Consider an arbitrary function $g(X)$ then, we have

$$E\left[(X - E(X|Y))^2\right] \leq E\left[(X - g(Y))^2\right],$$

  that is the expected square "distance" between $g(Y)$ and X is minimized for $g(Y) = E(X|Y)$.
- The proof is valid for both discrete and continuous r.v.

Proof

- We have

$$
\begin{aligned}
& E\left[\left(X - g\left(Y\right)\right)^2 \middle| Y = y\right] \\
&= E\left[\left(X - E\left(X|\,Y\right) + E\left(X|\,Y\right) - g\left(Y\right)\right)^2 \middle| Y = y\right] \\
&= E\left[\left(X - E\left(X|\,Y\right)\right)^2 \middle| Y = y\right] + E\left[\left(E\left(X|\,Y\right) - g\left(Y\right)\right)^2 \middle| Y = y\right] \\
&+ 2E\left[\left(X - E\left(X|\,Y\right)\right)\left(E\left(X|\,Y\right) - g\left(Y\right)\right)\middle| Y = y\right]
\end{aligned}
$$

- We have

$$
\begin{aligned}
& E\left[\left(X - E\left(X|\,Y\right)\right)\left(E\left(X|\,Y\right) - g\left(Y\right)\right)\middle| Y = y\right] \\
= \; & \underbrace{E\left[\left(X - E\left(X|\,Y\right)\right)\middle| Y = y\right]}_{=0}\left(E\left(X|\,Y = y\right) - g\left(y\right)\right)
\end{aligned}
$$

Proof

- Hence

$$E\left[\left.(X-g(Y))^2\right|Y=y\right] = E\left[\left.(X-E(X|Y))^2\right|Y=y\right] + \underbrace{E\left[\left.(E(X|Y)-g(Y))^2\right|Y=y\right]}_{\geq 0}$$

- We can conclude that

$$E\left[\left.(X-g(Y))^2\right|Y=y\right] \geq E\left[\left.(X-E(X|Y))^2\right|Y=y\right].$$

- Now by taking the expectation on both sides with respect to Y, we obtain

$$E\left[(X-g(Y))^2\right] \geq E\left[(X-E(X|Y))^2\right].$$