Conditional probability and Bayesian estimation

x, y are r.v. with joint distribution $p_{xy}(x,y)$

$x|y$ is a r.v. with distribution $p(x|y) = \frac{p_{xy}(x,y)}{p_y(y)}$

This describes scenarios where we might be able to observe y, to make more precise statements for x.

Note: If x, y are independent $p(x|y) = p_x(x)$
→ there is no information in y about x.

Conditional expectation:

$$E[x|y] = \int x \, p(x|y) \, dx \quad \text{is a r.v.}$$

$E[x|y]$ is best approximation of x

$$E[(x - E[x|y])^2] \leq E[(x - g(y))^2] \text{ for any } g.$$

Why?

$$E[(x-g(y))^2] = \iint (x-g(y))^2 p_{xy}(x,y) \, dx \, dy$$

$$= \iint (x-g(y))^2 p(x|y) \, p(y) \, dx \, dy$$

$$= \iint \underbrace{(x^2 - 2x g(y) + g(y)^2) \, p(x|y)}_{\text{minimize this}} \, p(y) \, dy \, dx$$

$$\int (x^2 - 2x g(y) + g(y)^2) \, p(x|y) \, dx = \int x^2 p(x|y) \, dx - 2 g(y) E[x|y] + g(y)^2$$

$$\Rightarrow g(y) = E[x|y]$$

Bayesian estimation

$$p(x|y) = \frac{p(x,y)}{p(y)} \quad , \quad p(y|x) = \frac{p(x,y)}{p(x)}$$

$$p(x|y) = p(y|x)\,\frac{p(x)}{p(y)}$$

$$\underset{\uparrow}{p(x|y)} \propto \underset{\uparrow}{p(x)}\,\underset{\uparrow}{p(y|x)}$$

Posterior prior likelihood

Suppose x is unknown with prior probability $p(x)$

Suppose y is a measurement of x:

$$y = h(x) + \eta$$

$\uparrow$ r.v. with known distribution

$$p(y|x) = p_\eta(y - h(x))$$

$$\Rightarrow p(x|y) \propto p(x)\, p_\eta(y - h(x))$$

↳ we can compute posterior distribution based on prior & likelihood.

Example: $p(x) \sim N(0,1)$

$$y = x + \eta, \quad \eta \sim N(0,1)$$

$$p(x|y) \propto \exp(-\tfrac{1}{2}x^2)\exp(-\tfrac{1}{2}(y-x)^2)$$

$$\propto \exp(-\tfrac{1}{2}\underbrace{(x^2+(y-x)^2)}_{F(x)})$$

$$\left.\begin{aligned} F(x) &= \min F(x) + \tfrac{1}{2}(x-\mu)^2 F''(\mu) \\ \mu &= \arg\min F(x) \end{aligned}\right\} \text{Taylor expansion of } F \text{ around minimize}$$

$$F'(x) = x + (x-y) = 0 \Rightarrow \mu = \tfrac{1}{2}y$$

$$F''(x) = 2$$

$$F(x) = \text{Const} + \tfrac{1}{2}(x-\tfrac{1}{2}y)^2\cancel{2}\cdot 2$$

$$\hookrightarrow p(x|y) \propto \exp\left(\tfrac{1}{2}\frac{(x-\tfrac{1}{2}y)^2}{\tfrac{1}{2}}\right) = N(\tfrac{1}{2}y, \tfrac{1}{2})$$

$E[x|y]$ — variance is reduced

The main part of this class will be about how to deal with complex posterior distributions.

Some useful tricks

(i) $p(x) \propto \exp(-F(x))$

$F(x)$ is quadratic

$p(x) = \mathcal{N}(\mu, \Sigma)$ where $\mu = \arg\min F$

$$\Sigma = \left(\frac{\partial^2 F}{\partial x^2}\Big|_{x=\mu}\right)^{-1}$$

(ii) $p(x|y) \propto p(x)\,p(y|x)$

$p(x|y,z) = p(x|y)$ if x, z are independent

$p(x,y|z) = p(x|y,z)\,p(y|z)$

Basic Monte Carlo

η_i are iid $\left(E[\eta_i] = \hat{\eta} \right.$, $\left. Var(\eta_i) = \sigma^2 \right)$ $\quad p_i(\eta_i) = p(\eta_i)$

<u>Define:</u> $\eta = \frac{1}{n} \sum_{i=1}^{n} \eta_i$

Recall Chebyshev: $P(|\cancel{\eta} - E[x]| \geq k\sigma) \leq \frac{Var(x)}{(k\sigma)^2}$

$$E[\eta] = \frac{1}{n} \sum E[\eta_i] = \hat{\eta}$$

$$Var(\eta) = Var\left(\frac{1}{n} \sum \eta_i\right) = \frac{1}{n^2} \sum Var(\eta_i) = \frac{1}{n} \sigma^2$$

$$\rightsquigarrow \underbrace{P\left(|\eta - E[\eta]| \geq k \cdot \frac{\sigma}{\sqrt{n}}\right)}_{} \leq \frac{1}{k^2}$$

$\rightsquigarrow$ The larger n is, the smaller is the error $|\eta - E[\eta]|$.

$\Rightarrow$ We can compute $E[\eta]$ by repeatedly doing an independent experiment and averaging!

$\Rightarrow$ Error decreases as $n^{-1/2}$. $\rightsquigarrow$ the more samples / experiments, the better.

<u>Basic idea of MC:</u>

$$E[f(x)] \approx \frac{1}{n} \sum f(x_i) \qquad x_i \sim p(x)$$

<u>Examples:</u> 1) $E[x] = \frac{1}{n} \sum x_i$

2) $Var(x) = \frac{1}{n} \sum (x_i - \mu)^2$

Problem: we don't know μ!

$$\mathrm{Var}(x) \approx \frac{1}{n} \sum (x_i - \bar{x})^2 \quad , \quad \bar{x} = \frac{1}{n} \sum x_i .$$

This works, but it is not great.

<u>Def.:</u> A "statistic" is a fctn of a sample.

Example: $\bar{x} = \frac{1}{n} \sum x_i$

<u>Def:</u> A statistic $\hat{\theta}$ is unbiased if $E[\theta] = E[\hat{\theta}]$

One can show (see HW):

$\frac{1}{n} \sum (x - \bar{x})^2$ is not unbiased

but $\frac{1}{n-1} \sum (x - \bar{x})^2$ is unbiased.

For immediate use:

$$E[x] \approx \frac{1}{n} \sum x = \bar{x}$$

$$\mathrm{Cov}(x) \approx \frac{1}{n-1} \sum (x - \bar{x})(x - \bar{x})^T$$

How do we draw samples of a r.v. using a computer?

There are basic algorithms to draw samples of uniform r.v.
Using these, we can construct Gaussians:

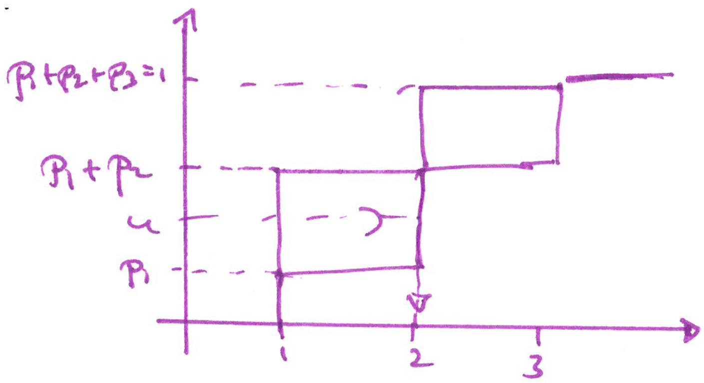
$$y_1 = \sqrt{-2\sigma^2 \log x_1}\, \cos(2\pi x_2) \qquad x_1\, x_2 \sim U[0,1]$$
$$y_2 = \sqrt{-2\sigma^2 \log x_1}\, \sin(2\pi x_2)$$

$$y_1, y_2 \sim N(0, \sigma^2) \qquad \text{"Box-Müller"}$$

In general:

1) Construct $F(x)$ (PDF)
2) Solve $F(x) = u$ for uniform u.

Why? Approximate $F(x)$ by step fcts



$u < p_1$ with prob p_1

$p_1 < u < p_1 + p_2$ with prob p_2

$p_1+p_2 < u < p_1+p_2+p_3$ with prob p_3

This is deceivingly simple! (i) $F(x)$ requires high-D integration

(ii) Solving $F(x)=u$ is not easy
(1 eqn in n variables)

A large part of this class is devoted to drawing samples from complicated distributions.